United States
Environmental Protection
Agency

Biological Indicator Variability and Stream Monitoring Program Integration: A Maryland Case Study

EPA/903R-02/008
October 2002

Biological Indicator Variability and Stream Monitoring Program Integration: A Maryland Case Study

Submitted to:

Technology Planning & Management Corporation
Mill Wharf Plaza, Suite 208
Scituate, MA 02066

Prepared for:

Wayne S. Davis
U.S. Environmental Protection Agency
Office of Environmental Information
Environmental Analysis Division
Mid-Atlantic Integrated Assessment Team
701 Mapes Road
Ft. Meade, MD 20755-5350

Prepared by:

Nancy Roth
Jon Vølstad
Ginny Mercurio
Mark Southerland

Versar, Inc.
9200 Rumsey Road
Columbia, MD 21045

Printed on chlorine free 100% recycled paper with
100% post-consumer fiber using vegetable based ink.

FOREWORD

This report was prepared by Versar, Inc., with support from the U.S. Environmental Protection Agency (EPA contract number 68-D-98-002 with Technology Planning and Management Corporation, subcontract 98-6026-98-005-Versar, Work Assignment 23 to Versar, Inc.). The mention of trade names, commercial products, or organizations does not imply endorsement by the U.S. Government.

Additional funding was provided by Maryland Department of Natural Resources, Power Plant Research Program (Biology Integrator, Contract No. PR-96-055-001 to Versar, Inc.).

The preferred citation for this report is:

Roth, N.E., J.H. Vølstad, G. Mercurio, and M.T. Southerland. 2001. Biological Indicator Variability and Stream Monitoring Program Integration: A Maryland Case Study.. Prepared by Versar, Inc., Columbia, MD, for U.S. Environmental Protection Agency, Office of Environmental Information and the Mid-Atlantic Integrated Assessment Program.

ACKNOWLEDGEMENTS

This project was a cooperative effort made possible by the important contributions of Ronald Klauda and Paul Kazyak of the Maryland Department of Natural Resources' Maryland Biological Stream Survey; Richard Eskin of the Maryland Department of the Environment, Keith Van Ness of the Montgomery County Department of Environmental Protection's Biological Monitoring Program, Wayne Davis and Ronald Shafer of U.S. EPA's Office of Environmental Information, and Nagaraj Neerchal of the University of Maryland Baltimore County's Department of Mathematics and Statistics. The authors also wish to thank Allison Brindley at Versar, Inc., for geographic information system support and several reviewers for their constructive comments.

EXECUTIVE SUMMARY

VARIABILITY OF BIOLOGICAL INDICATORS

A regulatory decision-making framework is currently being developed by the Maryland Department of the Environment (MDE) for listing watersheds as impaired (Clean Water Act, Section 303 (d)), using the Maryland Department of Natural Resources' (DNR) Maryland Biological Stream Survey (MBSS) Indices of Biotic Integrity (IBI) scores for fish and benthic macroinvertebrates. The MBSS uses both fish and benthic macroinvertebrate IBIs based on a suite of community-based metrics to characterize the health of freshwater streams statewide. In this report, we use a model-based approach to quantify the uncertainty around biological indicators at individual sites and we discuss how such uncertainty can be taken into account in the biocriteria framework. Key findings are summarized below:

> The MBSS conducts replicate <u>benthic sampling</u> at a random subset of stream segments each sampling year. We used data from 27 sites to assess the level of agreement between replicate samples and the average variability in benthic IBI scores within stream segments.
>
> Because it is not possible to collect replicate <u>electrofishing samples</u> within a stream segment and no data were available from adjacent segments, we used MBSS data from two or more sites sampled in the same reach within the same year as a surrogate. Analyses were restricted to reaches with pairs of sites less than 1.0 km apart and with similar land uses, water chemistry, and physical habitat. Replicate samples from 53 reaches were used to estimate fish IBI variability for the biocriteria framework.
>
> The average coefficient of variation (cv) for replicates was estimated at 8% for both fish and benthic macroinvertebrate IBI scores, suggesting homogeneous fish and benthic communities at a local spatial scale.
>
> We also measured the reliability of fish and benthic IBI scores at individual sites, i.e., the extent to which a survey of a watershed will provide the same results with repeated measurement at the same stream segments. Results from the 27 sites with replicate samples suggest that the MBSS sampling protocol results in reliable IBI scores.

INTEGRATION OF THE MBSS AND COUNTY MONITORING PROGRAMS

Several counties are conducting stream monitoring programs at a local scale, using field sampling protocols similar to those used by the MBSS. Using the Montgomery County stream monitoring program, we outlined how the statewide MBSS can be integrated with local scale stream monitoring programs to improve the estimation of stream condition in local areas and to provide consistent and reliable statements to the public. This study is based on information from the 1995-97 MBSS, 1994 MBSS Demonstration study, 1993 Pilot Study, and data from a field methods comparison study conducted jointly by MBSS and Montgomery County in 1997. Key findings are summarized below:

> The MBSS and Montgomery County monitoring programs have important differences in objectives. A primary goal of the MBSS is to estimate the status of streams, both statewide and at the Maryland 8-digit watershed level (a unit smaller than the USGS 8-digit cataloging unit). Montgomery County is primarily interested in assessing the status of streams in local areas (e.g., a sub-watershed or finer spatial scale) and in monitoring conditions downstream of specific developed areas.
>
> Key goals for program integration are to develop consistent statements to the public about stream conditions within Montgomery County, increase accuracy in estimates of stream condition in local areas, and reduce costs of the sampling programs by eliminating duplication of effort.
>
> Effective program integration requires extensive information beyond the basic monitoring data, including GIS files of streams, watershed boundaries, definition of the geographic strata used in site selection and indicator development (e.g., ecoregions, subwatersheds, soil types, or other regional strata); similar training of field personnel; field sampling manuals and field data sheets; and procedures for calculating the IBIs of both programs.
>
> Because objectives differ between programs, maps of different scales are used to plan the field sampling effort. GIS analyses of the 1:24,000 map used by the Montgomery County and the 1:100,000 map used in the second round of MBSS revealed a large overlap (202 stream miles), but a substantial number of streams were only found on the 1:24,000 map (120 stream miles). Only 7 miles of streams were exclusive to the 1:100,000 map. The 1:24,000 map thus improves stream coverage, particularly in the subset of small headwaters. The map scale also influences the stream order designation.
>
> The survey designs for the MBSS and County program support area estimates of stream condition, but differ at several levels. In the MBSS, a stratified random sample of stream segments is selected within watersheds. Montgomery County uses both targeted and probability-based sampling of reaches in a watershed to support different management needs, with random site selection within reaches.

Several differences in field sampling protocols exist between the two programs. Montgomery County has used three electrofishing passes, while the MBSS uses two. No significant differences were found between fish IBI scores based on two versus three passes. Montgomery County samples benthic organisms with two kick net samples in riffle habitat only, identifies up to 200 benthic organisms in the lab, and only identifies oligochaetes and chironomids to family; MBSS samples with a D-net in a variety of habitats (primarily riffles) using 20 jabs, identifies up to 100 organisms in the lab, but mounts and identifies oligochaetes and chironomids to genus or lowest possible taxon. Analyses in this report suggest that these differences can have a significant effect on IBI scores. Maryland DNR and Montgomery County have each developed fish and benthic IBIs that differ from one another. Benthic IBIs from both programs were compared at sites sampled jointly by the two programs, but the results were inconclusive, owing to the small number of sites (12).

A conceptual approach for obtaining integrated estimates of stream condition for the overlapping streams was developed and is outlined in this report.

We recommend that a field experiment be conducted to address the unresolved issues that may affect benthic IBI comparability. To this end, we present in this report a study design for a pilot project to assess these effects.

CONTENTS

TABLES

FIGURES

Figure No. **Page**

FIGURES (CONT'D)

Figure No. **Page**

FIGURES (CONT'D)

Figure No. **Page**

1. INTRODUCTION

The Maryland Biological Stream Survey (MBSS) is a long-term program conducted by Maryland Department of Natural Resources (DNR) to assess the condition of the state's freshwater, nontidal streams. Major accomplishments of the first MBSS sampling round (1995-1997) included sampling nearly 1000 sites statewide, development of ecological indicators of stream conditions, and completion of a comprehensive assessment of stream conditions, including estimates statewide and within major drainage basins. Results are currently being used to support Maryland's development of biological criteria and to evaluate conditions at finer watershed scales. To meet the State's growing need for finer-scale assessments, a modified study design was adopted for the Survey's second round (2000-2004) to provide more precise assessments at the Maryland 8-digit watershed scale, in addition to basin and statewide estimates.

As DNR embarked on this second round of statewide sampling, new issues of interest to resource managers were identified for investigation. One was the need for further analyses to determine the best approach for using nontidal stream monitoring data and indicators to support the development of biological criteria. In particular, quantifying the variability of Index of Biotic Integrity (IBI) scores is important to establishing thresholds for determining biological impairment under the State's interim biological criteria framework (MDE 2000). Another area of interest was determining the best approach to integrating county, state, and other monitoring programs. There is a need for cost-effective integration of MBSS with other stream monitoring programs in Maryland, particularly with the growth of county and local monitoring efforts spurred by local concerns and NPDES stormwater permit requirements. Both issues reflect the increasing use of biological data for a variety of purposes, including watershed management at scales finer than those previously considered by MBSS. Integration of county and other monitoring data offers the opportunity to supplement the statewide coverage of MBSS with more local-scale information, thus providing more data for assessing small watersheds or diagnosing problems at specific sites.

This report documents recent work to address these issues. Specific topics for investigation were identified through discussions with Maryland DNR, Maryland Department of the Environment (MDE), Montgomery County Department of Environmental Protection (DEP), and U.S. Environmental Protection Agency (EPA). Analyses reported here are intended to support programs in all four agencies related to the assessment and management of stream resources in Maryland, with potential applications to other states. Chapter 2 of this report covers IBI analyses in support of biological criteria development. Chapter 3 presents general guidelines for stream monitoring program integration, using integration of the MBSS with the Montgomery County, Maryland, Biological Monitoring Program as an example.

2. EVALUATION OF RELIABILITY AND PRECISION OF MBSS IBI SCORES

2.1 BACKGROUND

A regulatory decision-making framework is currently being developed by MDE for listing watersheds as impaired (303(d) list), based on information from MBSS IBI scores for fish and benthos. Maryland divides its waters into 138 8-digit watersheds, a scale finer than the USGS 8-digit hydrologic unit codes (Table 2-1). For some purposes, these watersheds are further divided into 12-digit subwatersheds. For Maryland 8-digit watersheds with MBSS samples from ten or more representative sites (i.e., 75-m stream segments), the proposed interim biocriteria framework would list watersheds as impaired by comparing mean IBI scores and confidence levels with a threshold value that flags degraded watersheds. For 8-digit watersheds with less than 10 representative samples, 12-digit subwatersheds that have one or more sites with IBI scores below a threshold value would also be listed.

Table 2-1. Comparison of USGS and Maryland hydrologic units.

	USGS 8-digit cataloging unit	MD 8-digit watershed	MD 12-digit subwatershed
Number in Maryland	20	138	1066
Average size in Maryland (approx.)	500 sq. mi.	75 sq. mi.	8 sq. mi.

A primary objective of our study was to derive quantitative values for the uncertainty in single-site IBI scores, to assist in developing appropriate criteria for listing 12-digit subwatersheds. We assess the uncertainty around biological indicators at individual sites and also discuss the validity of extrapolating results from individual sites to larger areas. We also compare the within-site variability to larger area variability. Only limited data were available for assessing the within-site variability in IBI scores. It is expected that in the future, as more data are collected, estimates of IBI variability will be more accurate.

The primary source of data for developing and implementing the biocriteria framework is the statewide MBSS (Klauda et al. 1998). The first round of the MBSS, conducted from 1995 to 1997, was primarily designed to provide reliable information on stream conditions for Maryland's major basins (Roth et al. 1999). Approximately 300, non-overlapping 75-m stream segments (sites) were sampled each year from non-tidal streams of first, second, and third order. The streams were defined using a 1:250,000-scale map and the segments were randomly selected using a lattice sampling approach that ensured coverage of the entire state over the three-year cycle (Klauda et al. 1998, Heimbuch et al. 1999). The MBSS uses both fish and benthic macroinvertebrate indices (Roth et al. 2000, Stribling et al. 1998) based on a suite of community-based metrics to characterize the health of freshwater streams.

The first round of the MBSS was not designed to provide estimates of stream condition for individual 8-digit watersheds. Instead, samples were selected in each of 17 larger drainage basins across the entire state (approximately the size of USGS 8-digit hydrologic units). However, estimates of stream condition within Maryland 8-digit watersheds can be obtained by post-stratifying basins with adequate sampling coverage. The second round of the MBSS, beginning in 2000, was designed to provide reliable estimates of stream condition for all 8-digit watersheds during a five-year cycle. A minimum of 10 random samples will be collected within each 8-digit watershed or a combination of small 8-digit watersheds.

2.2 DATA SOURCES

This study is based on information from the MBSS 1995-1997 (Roth et al. 1999, Klauda et al. 1998), the 1994 MBSS Demonstration study (Vølstad et al. 1996), and the 1993 Pilot Study (Vølstad et al. 1995). Data from a field methods comparison study conducted jointly by MBSS and Montgomery County in 1997 is analyzed in Chapter 3. In the MBSS, benthic macroinvertebrates are collected to provide a qualitative description of the community composition at each 75-meter stream segment (Kazyak 2000). Composite sampling, defined as the pooling of field samples prior to laboratory studies, is used to enhance the accuracy of estimated parameters meant to characterize the benthic communities in a stream segment. In the MBSS, a total of 20 plots are sampled within each stream segment using a 600 micron-mesh D-frame dipnet in riffles (if present) or in other representative habitat types such as snags, rootwads, or undercut banks. Benthic macroinvertebrates collected from these plots are pooled and, in the laboratory, a sub-sample of about 100 individuals is taken from this composite to estimate a benthic IBI (BIBI) score. This score is assigned to the 75-meter stream segment. Since the score is based on a composite sample of organisms from 20 small plots (2 m^2 total), and not on a census of the organisms within the 75-meter stream segment, the score will have an associated random error (sampling error). The 20-plot samples are likely to incorporate a significant portion of the variability in the benthic community at a site. However, because the jab samples are composited (pooled), the effect of between-plot variability cannot be assessed directly from the standard MBSS samples. Composite sampling and subsampling of the composite is applied to obtain a representative sample that closely approximates the information that would have been obtained from measuring the individual plots separately, but at reduced cost and effort (Patil et al. 1994).

To provide information on the uncertainty in BIBI scores at individual sites, the MBSS conducts replicate benthic sampling at a random subset of stream segments each sampling year. This within-segment uncertainty may affect the risk of misclassifying 12-digit subwatersheds as impaired based on BIBI scores at individual sites. If the within-segment variability in BIBI scores is significant, the precision of mean BIBI scores for 8-digit watersheds may also be affected. As such, the practice of conducting replicate sampling at a random subset of sites in the MBSS is an important component of the Survey, and will over time allow more accurate assessment of uncertainty and risks.

For fish sampling, the fish IBI (FIBI) for a single site based on the two-pass electrofishing sampling may also be inaccurate if the sampling is biased (e.g., if certain fish are collected out of proportion to the true occurrence). The fish sampling differs from the benthic sampling in that random replicates of electrofishing samples within a 75-m stream segment cannot be achieved because fish are removed from the stream segment in each pass, resulting in dependence between passes. The level of such bias in fish sampling can be assessed by conducting three or more passes at a representative subset of sites, and then comparing the FIBI scores based on two passes with the scores based on all passes. Small-scale variability in IBI scores can be evaluated by blocking two adjacent 75-m stream segments around random sites. The average variance in IBI scores for neighboring segments can be used to approximate the within site variability for FIBI scores, assuming habitat differences between adjacent segments are minimal.

As part of the quality assurance procedures for the 1995-1997 MBSS, two replicate benthic composite samples were collected at 27 randomly selected stream segments. We used these data to assess the level of agreement between replicate samples and the average variability in BIBI scores within stream segments. Benthic macroinvertebrates are generally sedentary and their spatial distribution within a segment is likely to be stable during the index period, barring intense stormflows. Replicate samples of fish assemblages at a local scale, even when conducted during the same day, are likely to exhibit variability in IBI scores because of fish movement and patchiness of fish communities. Sampling is conducted within an index period to minimize temporal effects, but short-term (e.g., daily) changes in distribution would affect the repeatability of future IBI scores at a segment level. Because it is not possible to collect replicate fish samples within a stream segment, we used as a surrogate MBSS data from two or more sites sampled in the same reach within the same year. An initial list of 100 reaches with two or more sites was filtered to identify reaches with pairs of sites less than 1.0 km apart and with similar land uses, water chemistry, and physical habitat. Replicate samples in the final group of 53 reaches were used to estimate FIBI variability for the biocriteria framework.

2.3 METHODS FOR EVALUATING UNCERTAINTY IN IBI SCORES

The classification success in the listing of watersheds as impaired depends on the uncertainty of estimated IBI scores for individual stream segments and of mean scores for watersheds. Measures of uncertainty can be broadly classified into accuracy and precision. In principle, accuracy refers to the size of deviations from the true mean [μ]; the precision is the degree of agreement between observations obtained by repeated application of the same sampling procedure (Cochran 1977). Data from a sample survey, and the values the data are used to calculate (estimates), have high precision if their error component is small. One measure of the random error, and thus of precision, is the error variance. This variance depends on several factors including the survey design, sample size, field sampling procedures, and subsampling in the laboratory (for benthos). The data and estimates calculated from the data may also involve a systematic error in addition to the random error. If the sampling gear and protocol results in the collection of only a portion of organisms present, or if some portion of the habitat is systematically under- or over-sampled, for example, the IBI scores would be biased. If the systematic error is large, deviations from the true value may be large, although precision may be high. The accuracy is said to be poor in such cases. If uncertainty can be quantified, decision makers will have a basis for evaluating the chances of incorrectly listing watersheds as impaired.

In evaluating the uncertainty of IBI scores, it is necessary to take into account the statistical survey design employed in the data collection. Two-stage sampling is employed in the MBSS to collect information on fish and benthos within a watershed. In the first stage, n stream segments are selected from a watershed by simple-random or stratified-random sampling; in the second step, subsamples of fish or benthos are collected within a stream segment. The sampling within 75-m stream segments for fish and benthos is different in principle.

We assume that benthic net samples within each stream segment are independent and representative. The principles of the study design for benthic field sampling within a stream segment is illustrated in Figure 2-1.

Assume that each stream segment i consists of a fixed number of habitat plots (M_i) that can be sampled by the net (e.g., riffle areas). Benthic samples are collected from m representative plots out of M_i plots within each of the n selected stream segments using a net. Thus nm benthic net samples are collected. It would be very costly to analyze all nm benthic samples in the laboratory. The composite sampling for benthos used in the MBSS and common in other stream sampling programs attempts to diminish this disadvantage: $m = 20$ plots from one stream segment are pooled into a composite sample. From each of the n stream segments, a fixed-count random subsample of organisms from the composite sample is analyzed in the laboratory. The composite sample involves a physical mixing of the m net samples. Such composite sampling/subsampling plans can greatly reduce the cost of laboratory analysis and can extract most of the information from the field samples (Edland and van Belle 1994, Boswell et al. 1988, Gilbert 1987). This approach makes the composite sample design very cost-effective while maintaining representativeness if properly conducted.

Sampling of fish within a stream segment is typically conducted by multipass electrofishing; the sequential passes are dependent. The multi-pass electrofishing sampling within a stream segment is illustrated in Figure 2-2.

The electrofishing can be considered a census of the stream segment; uncertainty in fish data and estimates within each stream segment is related to imperfections in the coverage of this census. When the combined passes fail to catch fish in their true proportion by species, the IBI for fish would be biased. To address this potential contribution to FIBI accuracy, we analyzed data from sites sampled by Montgomery County, using three electrofishing passes. We examined the effects of two versus three passes on species richness, abundance, and IBI scores (see section 3.6).

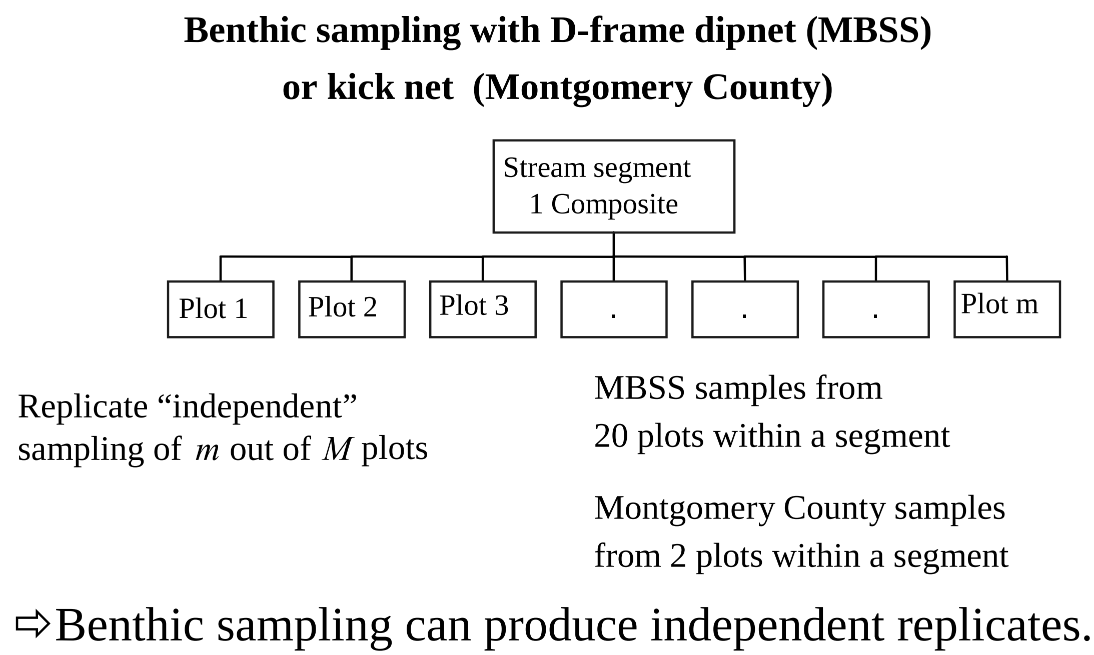


Figure 2-1. Schematic diagram of field sampling of benthos within a stream segment.

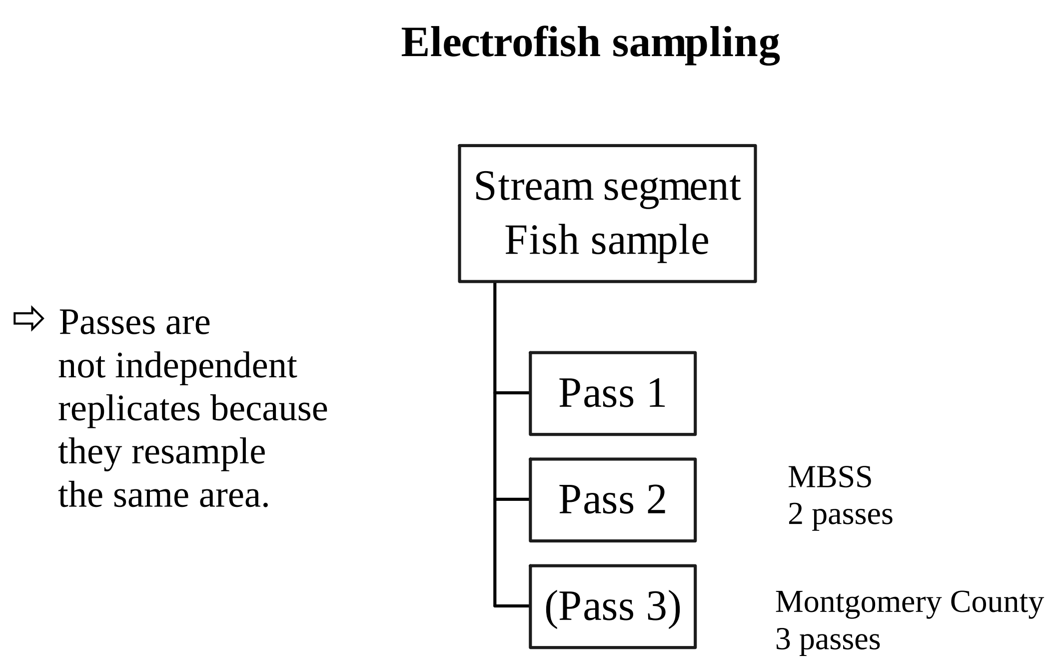


Figure 2-2. Schematic diagram of electrofishing sampling in a stream segment.

2.3.1 Measuring Reliability

For the biocriteria framework, it is important to know if IBI scores are an accurate representation of the stream condition at the sampled sites, as well as if repeated sampling of the same sites yields consistent, reliable results. In the context of the MBSS, we define reliability as the extent to which monitoring will provide the same results with repeated measurement at the same stream segments. We measured reliability by simple linear regression analysis, the intra-class correlation's coefficient, Cronbach's alpha, simple and weighted kappa, and the polychoric correlation coefficient.

The MBSS uses an IBI score from 1 to 5. Estimates of stream condition are classified into four categories based on the IBI scores: very poor ($1 \le IBI < 2$), poor ($2 \le IBI < 3$), fair ($3 \le IBI < 4$), and good ($IBI \ge 4$). The joint ordinal ratings for the replicate samples were displayed in a square table (e.g., upper panel of Figure 2-5). The main diagonal represents agreement for the ratings. This approach was also used to evaluate agreement in ratings of stream condition between the MBSS and the Montgomery County sampling programs (see section 3.6). We distinguish between measuring agreement and measuring association, because there can be strong association without strong agreement. For example, one sampling program may rate stream condition consistently one level higher than another program on an ordinal scale from very poor to good. If so, the strength of agreement is weak even though association is strong.

First, we conducted a linear regression analysis of raw IBI scores from replicate sampling of the same stream segments using the model,

$$IBI_2 = \alpha + \beta IBI_2$$

where IBI_2 is the score for the second sample and IBI_1 is the score for the first sample. The reliability of the IBI scores was assessed by the regression coefficients and the R^2. The regression plots also offered a simple visual means of determining whether the variability in IBI scores within stream segments tends to be greater for high or low mean scores.

Second, the intra-class correlation (ICC, or θ) was used to measure reliability of IBI scores based on replicate sampling within stream segments. The intra-class correlation may be conceptualized as the ratio of between-segment variance in IBI scores to total variance. An estimator for intra-class correlation is (Snedecor and Cochran 1980, p. 244)

$$\theta = \frac{s_b^2 - s_w^2}{s_b^2 + (m-1)s_w^2}$$

where

- s_b^2 is the ANOVA mean-square estimate of between-segment variance in IBI scores, reflecting the normal expectation that different stream segments will have different true scores on the rating variable.
- s_w^2 is the ANOVA mean-square estimate of within-segment variance in IBI scores, or error attributed to unreliability in rating the same segment based on replicate samples.
- m is the number of replicate samples within stream segments.

ICC will approach 1.0 when replicate samples within stream segments have equal IBI scores (i.e., when $s_w^2 = 0$).

The polychoric correlation (for ordered-category ratings) was used to measure agreement between categorical scores on an ordinal scale. The polychoric correlation is a maximum likelihood (ML) estimator for the correlation between two ordinal variables.

The strength of agreement between categorical scores was also measured by the *kappa* statistic (Agresti 1990). For independent replicate sampling in a randomly selected stream segment or reach, let π_{ij} denote the probability of classifying stream condition in the i th category based on the first sample, and in the j th category based on the second sample. Then

$$\Pi_0 = \sum \pi_{ii}$$

is the probability that the rating of stream condition based on the replicate samples agree. Perfect agreement means that the rating of stream condition is the same for both samples. If the ratings based on replicate samples were statistically independent, some agreement would still be expected purely by chance. The probability of agreement by chance is

$$\Pi_e = \sum \pi_{i+} \pi_{+i}$$

where π_{i+} is the probability of classifying the condition of the stream segment in the i th category based on the first sample, and π_{+i} is the probability of classifying the condition of the stream segment in the i th category based on the second sample. We used kappa as one technique for estimating how IBI scores from replicate samples within stream segments agree. The simple kappa,

$$\kappa = \frac{\Pi_0 - \Pi_e}{1 - \Pi_e}$$

is a measure of agreement that adjusts for the probability that some agreement will occur simply by chance. It has a scale ranging from zero (no better agreement than would be expected by chance) to 1 (perfect agreement). When categories are ordered (e.g., from poor to excellent), the seriousness of a disagreement depends on the difference between ratings. We therefore also calculated a weighted κ where agreement is higher for ratings that are closer together on an ordinal scale. We also calculated Cronbach's alpha to measure reliability (Cronbach 1951, Hughes et al. 1998).

2.3.2 Estimating Precision

Assume that a mean IBI ($\overline{x}$) is estimated from sampling in n randomly selected stream segments in the study area. An estimator of precision is the standard error, $SE(x) = \sqrt{\text{var}(\overline{x})}$. If the assumption that the estimate $\overline{x}$ is normally distributed around the corresponding population value holds, then lower and upper confidence limits for the mean IBI in the watershed are as follows (Cochran 1977):

$$\hat{\overline{X}}_L = \overline{x} - t\sqrt{\text{var}}(\overline{x}), \quad \hat{\overline{X}}_U = \overline{x} + t\sqrt{\text{var}}(\overline{x}).$$

The symbol *t* is the value of the normal deviate corresponding to the desired level of confidence and whether a one-sided or two-sided confidence interval is estimated. For one-sided confidence intervals, *t* is 1.28 for 90% confidence level, and 1.65 for 95% confidence level.

The variance in mean IBI has two sources: the first involves the variability in IBI scores between the *n* stream segments and the second involves the variability from sampling fish or benthos within stream segments. The total variance of the mean score can be expressed as

$$\text{var}\left(\overline{x}\right) = (1 - f_N)\frac{s_b^2}{n} + f_N(1 - f_M)\frac{s_w^2}{nm}$$

where

$f_N = \dfrac{n}{N}$ is the proportion of the N stream segments (or fraction of stream miles) actually sampled;

$f_M = \dfrac{m}{M_i}$ is the proportion of the M_i subunits in each stream segment actually sampled;

s_b^2 is an estimator of the variance in IBI score among all N stream segments; and

s_w^2 is an estimator of the variance of IBI scores among M plots within stream segments.

When the sampling fraction of stream segments is small, $f_N \approx 0$ and

$$\operatorname{var}(\bar{x}) \approx \frac{s_b^2}{n} \text{ and } SE(\bar{x}) \approx \frac{s_b}{\sqrt{n}} .$$

The expression for the variance of $\bar{x}$ in this case involves only the variability of the segment-level means and does not require estimation of the within-segment variability. This is because the within-segment variability is reflected in the variability of the segment mean IBIs. If a census is conducted within stream segments, $f_m = 1$ and the last component of the variance is zero. If multi-pass electrofishing within stream segments catches all fish, it would be a census, and $f_m = 1$.

If statements about IBI are made for individual stream segments, $f_n = 1$ and all uncertainty is expressed by the last component of the variance. When the m benthic samples collected within stream segments are pooled into one composite sample, it is not possible to estimate s_w^2. However, analysis of the 27 MBSS sites with replicate composite samples provides useful information on the expected variability in BIBI scores within sites. The estimated mean variability for sites with replicate benthic sampling can be used as an approximation for the variability at sites with only one benthic score. For FIBI scores, replicate samples within 75-m stream segments were not available. We used the estimated mean variability in FIBI scores from replicate samples within the same reaches (sites < 1 km apart and similar in character) to obtain an approximate estimate of the expected variance for individual sites. This estimate is likely to be conservative, because stream segments that are farther apart would be expected to yield IBI scores that are more variable than neighboring stream segments.

The precision of an estimated mean of sample values depends on the variability, or patchiness, of the population being studied and consistency of the field sampling. This natural variability between sampling units is usually dependent on the spatial scale of the survey. Replicate samples within a stream segment are expected to exhibit less variation than random samples within a watershed. We estimated the mean standard deviation ($\bar{s}_b$) for replicate sampling within stream segments, reaches, 12-digit subwatersheds, and 8-digit watersheds based on the 1994 MBSS and the 1995-1997 MBSS.

Another measure often used in describing the amount of natural variation in a population is the coefficient of variation: $CV = \sigma / \mu$. The CV is a relative index of variation that expresses the standard deviation of a parameter ( ) as a fraction, or sometimes as a percentage, of the mean ( ). For sample data on abundance, biomass, and species or taxa composition of fish or benthic organisms, the estimated mean and standard deviation often tend to be related (Seber 1973). MBSS IBI scores for fish and benthos have a range from 1 to 5 and thus the variance will be relatively small compared to typical abundance data for patchily distributed animals. We investigated whether IBI scores also exhibited a relationship between the mean score and the variance of the scores. Patterns of spatial distribution determining the variation in sample values of IBI scores of fish and benthic communities are often complex and are influenced by a number of factors, including the spatial scale of the sampling, the size of the sampling unit, and the time of year. Patchiness, or clumping of organisms occurs at different scales and is influenced by environmental factors. Instream habitat features play an important role in the distribution of biota. The MBSS is designed to reduce the effects of these factors in the estimation of mean IBIs by (1) collecting representative samples over time and space, (2) using standardized sampling protocols, and (3) conducting the sampling within seasonal index periods.

In the evaluation of uncertainty in biological indicators, it is useful to have a measure of the amount of variation for key parameters that is relatively stable as their means vary. The *CV* is fairly robust to (i.e., buffered from) changes in the mean and is therefore a more useful measure of variation than the variance or standard deviation for assessing uncertainty and planning sample sizes in future surveys. An estimator of the *CV* from sample values is $cv = s / \bar{x}$ where s is the estimated standard deviation. The *cv* is a measure of the degree of patchiness (or clumping) of the population being sampled.

We used MBSS data from 1995-1997 to estimate the average standard deviation and *cv* for fish and BIBIs. In MBSS 1995-1997 two replicate benthic samples were collected from each of 27 randomly selected stream segments as part of the quality control. These data were used to estimate the average variability in IBI scores between repeated samples at the same site.

In the evaluation of uncertainty in FIBI and BIBI scores, it is useful to analyze the MBSS data at different spatial scales. Replicate benthic samples within individual stream sites, or from neighboring sites, are likely to be more similar than scores from random stream locations in a larger geographic area, because the environment generally is less variable in smaller areas. Random sampling from a collection of streams (e.g., within a watershed) provides representative information on the mean IBI score for the entire study area. The results, however, are only applicable to the streams actually included in the sampling frame. Similarly, random sites within a reach would produce unbiased estimates of mean condition for that reach. Individual IBI scores, in contrast, only represent the stream segment sampled. It is generally not valid to extrapolate sample information from a single site to a larger area.

In general, uncertainty in mean IBI scores depends both on (1) the spatial and temporal variation in the communities being studied and (2) the study design and sample sizes of the sampling program for collecting information on these indicators.

We used post-stratification to estimate average variability in 1995-97 IBI scores for random samples within reaches, 12-digit subwatersheds, and 8-digit watersheds. The first round of MBSS involved representative sampling of streams throughout the state. We also analyzed FIBI scores from replicate samples within reaches using MBSS 1994 data. In the 1994 demonstration study, 60 reaches from 6 major basins had two or more electrofishing samples; 54 reaches had samples less than 1 km apart. BIBI scores were not available for the 1994 data.

The expected precision in mean IBI scores depends on the variability of the assemblages being sampled and on the actual sample size. A practical measure of precision for use in the evaluation of uncertainty of biological indicators is to calculate the relative standard error of the mean IBI estimates (Jessen 1978). Because the *cv* appears to be related to the mean IBI, the relative standard error will be a more stable measure of uncertainty when comparing IBIs across streams with different conditions. The relative standard error is defined as

$$RSE(\bar{x}) \quad \frac{SE(\bar{x})}{\bar{x}} \quad \frac{cv}{\sqrt{n}}.$$

Assuming random sampling, the relative standard error of an estimated mean thus depends on the population *cv* and the sample size n. Since the *CV* is more stable with respect to changes in the mean, the RSE will be less variable than the standard error. A rough assessment of the average uncertainty around BIBI scores at individual stream segments can be based on the average *cv* for stream segments with replicate benthic sampling. However, because of the patchiness of benthic communities, an average measure from multiple representative sites can be misleading for any individual site. However, an average measure of uncertainty is indeed useful for assessing the general risk of misclassifying watersheds.

We may desire that the relative standard error be within a certain percentage of the mean of a parameter, regardless of the size of the mean, rather than specifying a fixed value for the standard error. As such, the length of the confidence interval can also be specified relative to the mean. The relative length of the confidence limit for $\bar{x}$ can be expressed as (Jessen 1978)

$$\frac{e}{\bar{x}} \quad \frac{cv \quad t}{\sqrt{n}}.$$

Valid confidence intervals require that $\bar{x}$ be normally distributed. Violations of the normality assumption can result in erroneous estimates of confidence intervals.

2.3.3 Maryland's Interim Biocriteria Framework of Listing 12-digit Subwatersheds

Maryland proposes that 12-digit subwatersheds where one or more IBI scores fall below a threshold value be listed as impaired (MDE 2000). Recognizing the inherent uncertainty in using a single score to characterize a stream segment, the decision rule for determining impairment is not based on the score alone, but rather uses an interval estimate of the IBI that takes the uncertainty into account. An interval is constructed to include the "true" IBI score with a certain probability. We first estimate one-sided 90% confidence limits for mean IBI score ($\bar{X}$) based on relative standard errors:

$$CL_{Upper} = \bar{X} + t \times \bar{X} \times \frac{cv}{\sqrt{n}} \;, \qquad CL_{Lower} = \bar{X} - t \times \bar{x} \times \frac{cv}{\sqrt{n}}$$

For one replicate at a site, the mean value is the actual score and n = 1. The cv cannot be estimated for a sample size of 1. However, a model-based estimate can be based on the average cv at representative sites with replicate samples.

The proposed framework (MDE 2000) has these criteria:

- A 12-digit subwatershed is considered to be in good condition if the lower confidence bounds for both fish and BIBI are above or equal to 3 for all stream segments sampled.
- A 12-digit subwatershed is listed as impaired if the upper confidence limit for fish or BIBI is below 3 for one or more of the sampled stream segments.
- All 12-digit subwatersheds with MBSS samples that do not fall in either of these categories are indeterminate, suggesting that further sampling is required.

Samples of individual stream segments in the watersheds in need of further sampling fall into four groups:

I. Both IBI scores are above 3, but the lower confidence limit for one score is below 3;

II. Both IBI scores are above 3, but the lower confidence limit for both scores are below 3;

III. One IBI score is below 3, with upper confidence limit above 3, and one score is above 3, with lower confidence limit below 3;

IV. Both IBI scores are below 3, but the upper confidence limits are above 3.

A schematic outline of the classification framework proposed for 12-digit subwatersheds is in Figure 2-3.

2.4 RESULTS ON UNCERTAINTY IN IBI SCORES

2.4.1 Reliability

Measures of reliability in BIBI scores for replicate sampling within stream segments (based on analysis of data from MBSS 1995-97) are shown in Table 2-2, and Figures 2-4 and 2-5. The linear regression of score 2 against score 1 for duplicate samples shows a good fit ($R^2 = 0.72$) with a slope $\alpha = 1.01$ $(s.e. = 0.04)$. The increasing spread of the residuals around the fitted regression line indicates that the variability in IBI scores is higher in streams with good ratings than for streams with poor ratings (Figure 2-4). The categorical analyses show a relatively high level of agreement in stream ratings for replicate samples. For a majority of sites (> 70%), the replicate samples score in the same category, while the remaining sites have scores in neighboring categories (Figure 2-5). The high values for Cronbach's alpha, the weighted kappa, and the polychoric correlation (Table 2-2) indicate that the composite sampling for benthos results in highly reliable IBI scores at the segment level.

Table 2-2. Measures of reliability of BIBI scores for replicate samples within stream segments. Values in parentheses represent standard errors.

IBI	DATA	n	R^2	θ	*Cronbach's alpha*	κ_{simple}	κ_{wt}	ρ
Benthic	MBSS 1995-1997	27	0.72	0.85	0.92	0.57 (0.13)	0.70 (0.09)	0.91 (0.06)

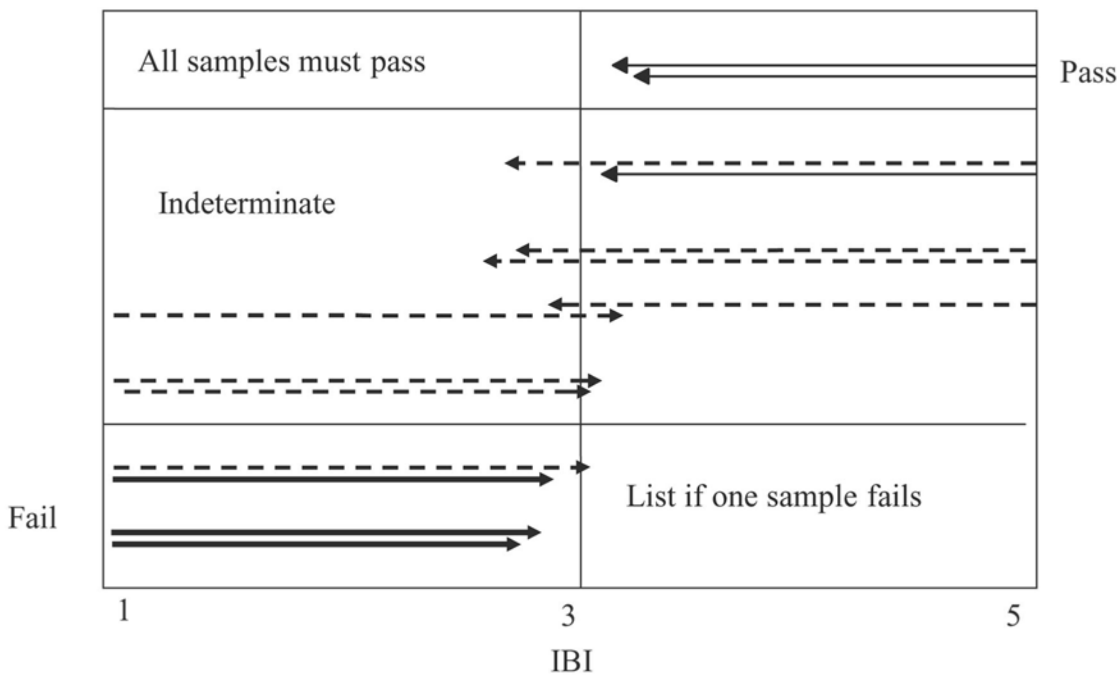


Figure 2-3. Schematic diagram illustrating Maryland's proposed biocriteria framework for identifying 12-digit subwatersheds as impaired. The system uses a fixed threshold value of 3 for IBI, and one-sided confidence interval estimates for the IBI scores. Pairs of arrows represent fish and BIBI results for an individual stream site. For scores less than 3, arrowheads represent the upper bound of the confidence interval; for scores less than 3, arrowheads represent the lower bound. Thick solid lined arrows signify scores that fail to meet criteria. Thin/solid lined arrows show scores that meet criteria. Dotted lined arrows indicate scores with confidence intervals that cross the threshold. A site is considered to "pass" if both IBI scores meet criteria, or "fail" if one or both scores fail to meet criteria. Within a 12-digit subwatershed, all samples (sites) must pass for the watershed to be considered passing. If one or more samples fail, the watershed may be listed as impaired.

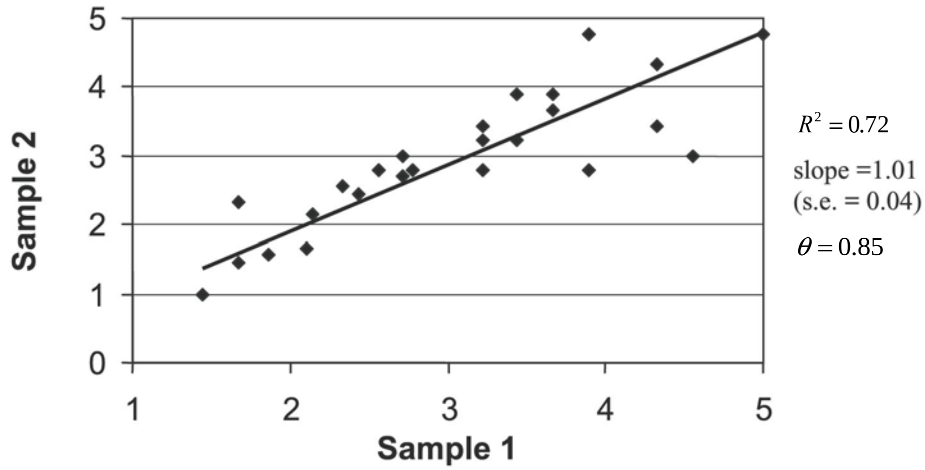


Figure 2-4. Regression analysis of BIBI scores for replicate samples, MBSS 1995-1997 (n = 27 sites).

MBSS sample 2					
	Good	0	0	**1**	**2**
	Fair	0	1	**8**	**2**
	Poor	1	**6**	2	0
	Very poor	**3**	1	0	0
		Very poor	Poor	Fair	Good
	MBSS sample 1				

Cronbach's alpha	κ_{simple}	κ_{wt}	ρ
0.92	0.57 (0.13)	0.70 (0.09)	0.91 (0.06)

Figure 2-5. Comparison of BIBI categorical classification of stream condition for replicate samples, MBSS 1995-1997. Values in parentheses represent standard errors.

2.4.2 Precision

Results for the replicate benthic sampling within 27 stream segments conducted as part of MBSS 1995-1997 are in Table 2-3. The average within-segment variability in BIBI scores increases with stream order, with cv's of 6%, 8%, and 11% for stream orders one to three (Table 2-4). The overall average cv of BIBI for replicate samples was 8% across stream order, i.e., the expected standard error is 8% of the mean score. The mean difference in BIBI scores between replicates (0.1) was not significantly different from 0 (p value $>$ 0.32; paired t-test). About 70% of the sites had a standard deviation of 0.25 or less (Figure 2-6).

Table 2-3. Summary data for replicate benthic composite samples at 27 randomly selected stream segments, conducted as part of MBSS 1995-1997. Score1 and Score2 represent rating categories assigned based on IBI (1 = very poor, 2 = poor, 3 = fair, 4 = good).

Stream order	IBI_1	IBI_2	ΔIBI	$\overline{x}_{dups}$	s_{dups}	cv_{dups}	Score1	Score2
2	2.71	2.71	0	2.71	0	0	2	2
2	3.22	3.44	-0.22	3.33	0.16	0.05	3	3
3	4.56	3	1.56	3.78	1.10	0.29	4	3
1	2.33	2.56	-0.23	2.45	0.16	0.07	2	2
3	3.89	2.78	1.11	3.34	0.79	0.24	3	2
1	1.67	2.33	-0.66	2	0.47	0.23	1	2
1	1.86	1.57	0.29	1.72	0.21	0.12	1	1
1	1.67	1.44	0.23	1.56	0.16	0.11	1	1
1	3.22	3.22	0	3.22	0	0	3	3
3	1.44	1	0.44	1.22	0.31	0.26	1	1
3	2.78	2.78	0	2.78	0	0	2	2
2	2.11	1.67	0.44	1.89	0.31	0.17	2	1
3	3.89	4.78	-0.89	4.34	0.63	0.15	3	4
1	3.22	3.22	0	3.22	0	0	3	3
2	5	4.78	0.22	4.89	0.16	0.03	4	4
1	3.67	3.67	0	3.67	0	0	3	3
2	4.33	3.44	0.89	3.89	0.63	0.16	4	3
3	4.33	4.33	0	4.33	0	0	4	4
2	3.67	3.89	-0.22	3.78	0.16	0.04	3	3
1	3.22	3.22	0	3.22	0	0	3	3
3	3.44	3.22	0.22	3.33	0.16	0.05	3	3
2	3.22	2.78	0.44	3	0.31	0.10	3	2
2	3.44	3.89	-0.45	3.67	0.32	0.09	3	3
3	2.43	2.43	0	2.43	0	0	2	2
1	2.14	2.14	0	2.14	0	0	2	2
1	2.71	3	-0.29	2.86	0.21	0.07	2	3
3	2.56	2.78	-0.22	2.67	0.16	0.06	2	2

Table 2-4. Means of BIBI ($\bar{x}$), standard deviation ($\bar{s}$) and coefficient of variation ($c\bar{v}$) for duplicate sampling within stream segments using data from MBSS 1995-1997.

		Stream order			
Metric	Statistic	1	2	3	All
	n	10	8	9	27
BIBI	$\bar{x}$	2.60	3.39	3.13	3.01
	$\bar{s}$	0.12	0.25	0.35	0.24
	$c\bar{v}$	0.06	0.08	0.11	0.08

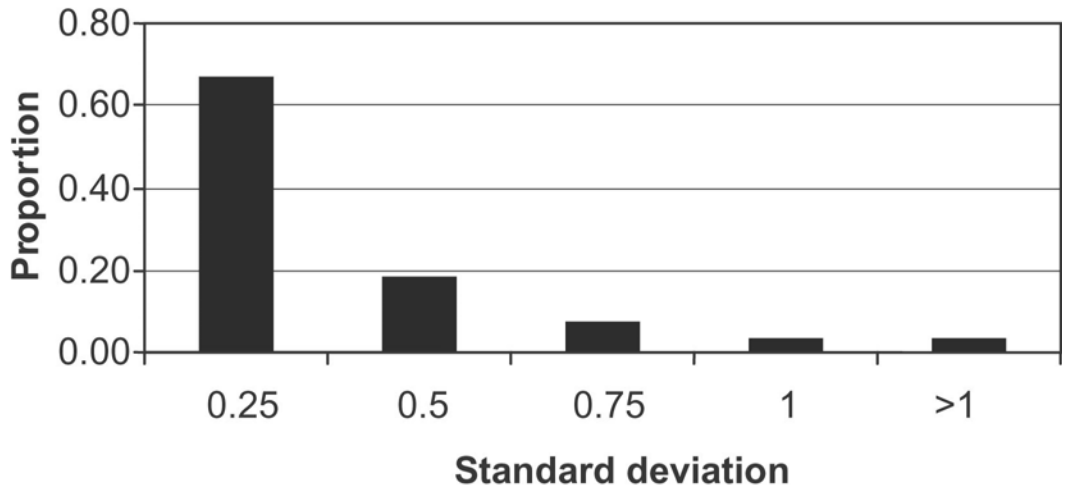


Figure 2-6. Distribution of standard deviation of BIBI scores for replicate samples within stream segments, MBSS 1995-1997.

The mean cv for FIBI scores within filtered reaches (sites < 1 km apart and of similar character) was also 8% (Table 2-5). The average variability for replicate FIBI scores is larger when data for all reaches are analyzed, with an average cv of 0.12 (Table 2-6). This increased variability is expected, because samples from segments that are farther apart were included.

Table 2-5. Means of FIBI ($\bar{x}$), standard deviation ($\bar{s}$) and coefficient of variation ($\overline{cv}$) within reaches by stream order using data from MBSS 1995-1997. The number of reaches (n) were filtered to remove those with sites > 1 km apart or sites with differing physical, chemical, or habitat features that indicated the presence of real differences in stressors.

		Stream order			
Metric	Statistic	1	2	3	All
	n	11	18	24	53
FIBI	$\bar{x}$	3.00	3.39	3.45	3.34
	$\bar{s}$	0.21	0.18	0.27	0.22
	$\overline{cv}$	0.06	0.06	0.09	0.08

Table 2-6. Means of FIBI ($\bar{x}$), standard deviation ($\bar{s}$) and coefficient of variation ($\overline{cv}$) within reaches by stream order using data from MBSS 1995-97. All reaches (including those > 1 km apart) were used in this analysis.

		Stream order			
Metric	Statistic	1	2	3	All
	n	29	33	39	100
FIBI	$\bar{x}$	3.11	3.44	3.49	3.35
	$\bar{s}$	0.44	0.34	0.34	0.37
	$\overline{cv}$	0.15	0.11	0.11	0.12

For comparison, mean standard deviation ($\bar{s}$) and coefficient of variation ($\overline{cv}$) of IBI ($\bar{x}$) were estimated for replicate sampling at several spatial scales. The MBSS 1995-1997 data included repeat sampling within stream segments (only benthic sampling), reaches, 12-digit subwatersheds, and 8-digit watersheds. Average variability for the repeat samples for both fish and benthos generally increased with increasing spatial scale, with replicates within stream segments (benthos) or between segments in "filtered" reaches (fish) being least variable and replicates within 8-digit watersheds being most variable (Figures 2-7, 2-8, 2-9, 2-10). Reduced variability for FIBI scores within reaches was generally exhibited when the reach data were filtered to remove those with sites > 1 km apart or sites with differing physical, chemical, or habitat features that indicated the presence of real differences in stressors.

The expected precision of IBI scores in a study area depends on the spatial variability of the population being sampled and the sample size. Within stream segments, one composite sample is likely to characterize the benthic community quite accurately, with an expected relative standard error of 8%. For a study in a 12-digit subwatershed, in contrast, more than 10 samples would be required to achieve a relative standard error of 8%, on average, because of the increased spatial variability (Figure 2-11).

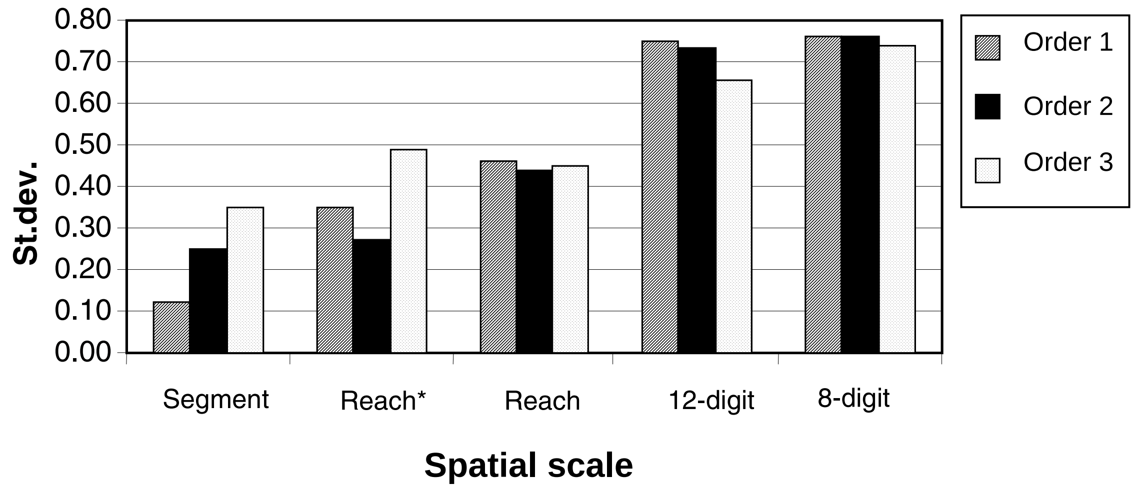


Figure 2-7. Mean standard deviation of replicate BIBI scores by stream order, for sampling at different spatial scales, using data from MBSS 1995-1997.
(*) Indicates reaches with pairs of sites less than 1.0 km apart, and with similar land uses.

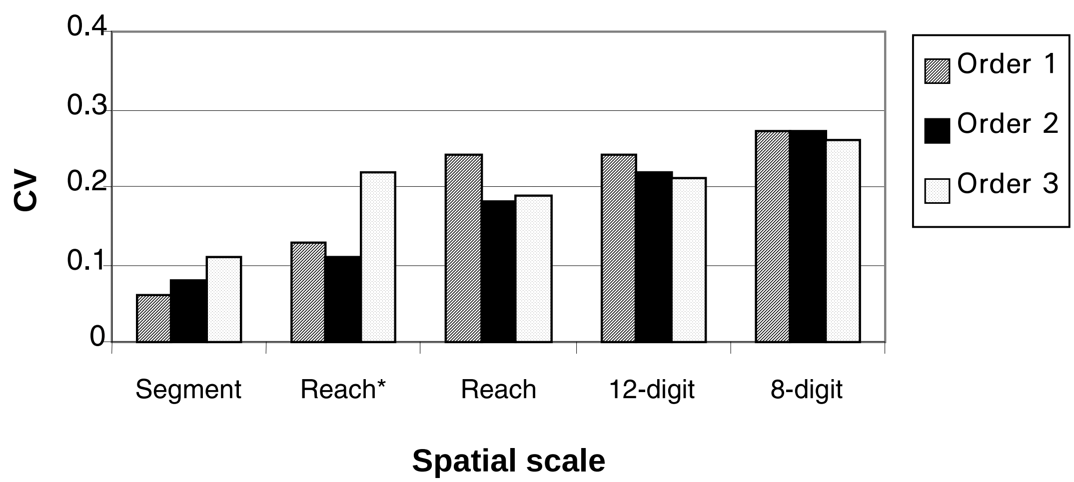


Figure 2-8. Mean coefficient of variation of replicate BIBI scores by stream order, for sampling at different spatial scales, using data from MBSS 1995-1997.
(*) Indicates reaches with pairs of sites less than 1.0 km apart, and with similar land uses.

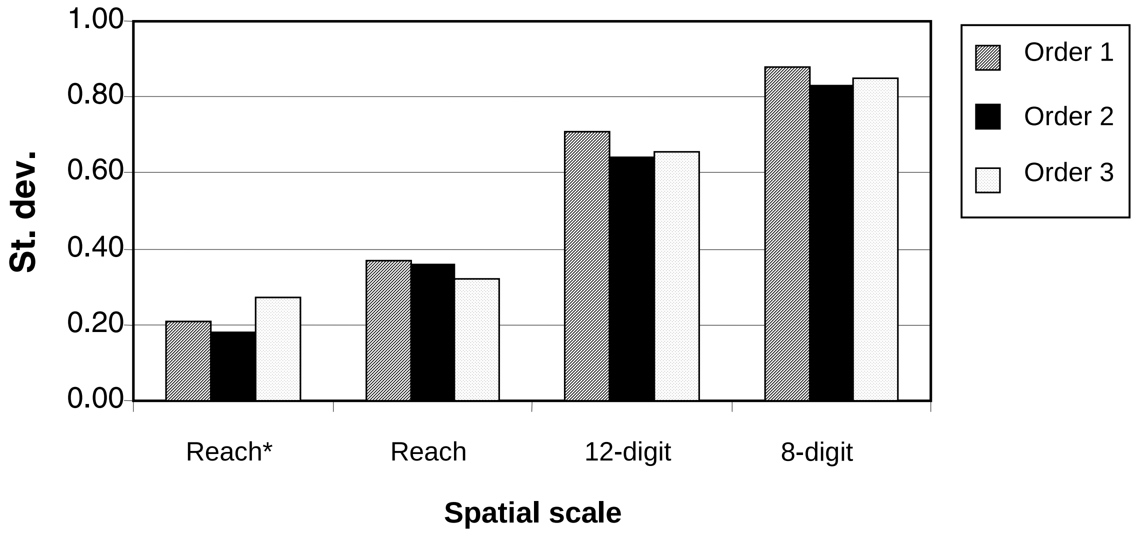


Figure 2-9. Mean standard deviation of replicate FIBI scores by stream order, for sampling at different spatial scales, using data from MBSS 1995-1997.
(*) Indicates reaches with pairs of sites less than 1.0 km apart, and with similar land uses.

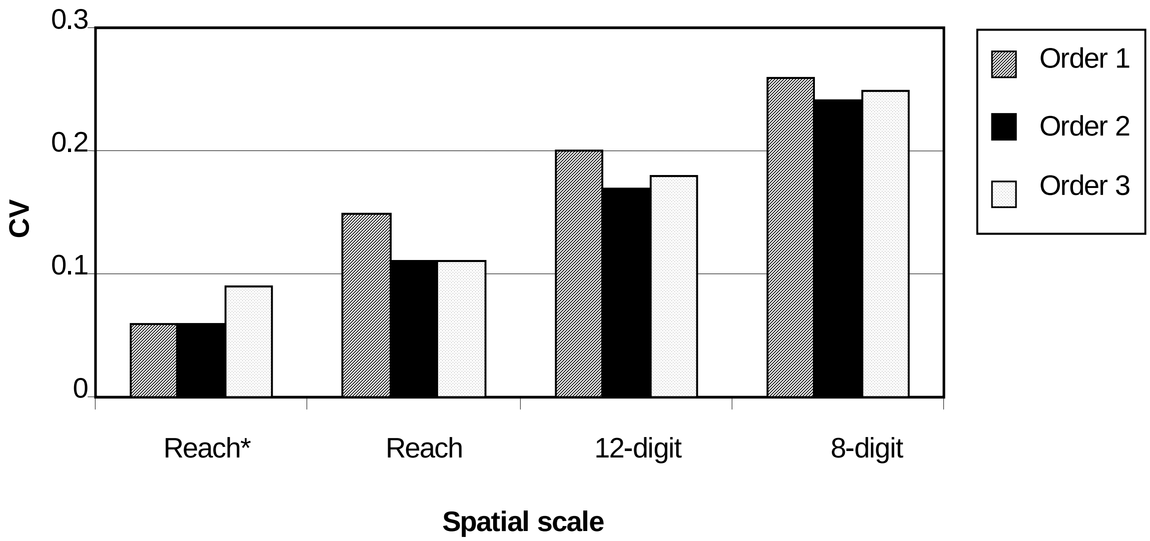


Figure 2-10. Mean coefficient of variation of replicate FIBI scores by stream order, for sampling at different spatial scales, using data from MBSS 1995-1997.
(*) Indicates reaches with pairs of sites less than 1.0 km apart, and with similar land uses.

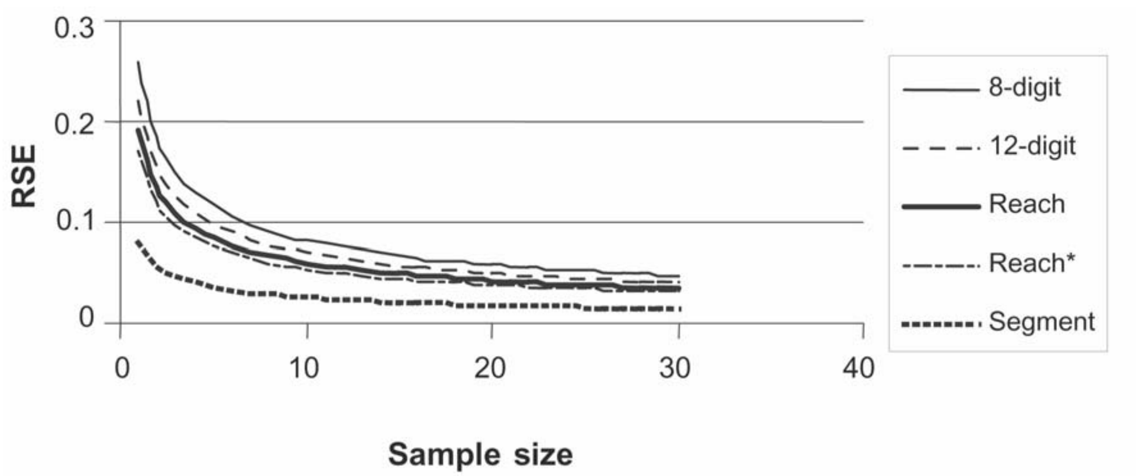


Figure 2-11. Relative standard error (RSE) as a function of sample size *n* at different spatial scales. (*) Indicates reaches with pairs of sites less than 1.0 km apart, and with similar land uses.

Variability did not appear to vary dramatically by stream order. For the FIBI, within-reach variability (measured either as standard deviation or *cv*) was slightly higher for first-order streams. This could result from the reduced average distance between sample locations for higher stream orders. The average length of 1st-order reaches sampled in MBSS 1995-1997 is 2.2 km, while 2nd- and 3rd- order reaches are 1.4 and 1.5 km long respectively. This effect was not observed when sites > 1 km apart were removed from the data set.

The coefficient of variation (*cv*) is a more stable measure of uncertainty than the standard deviation as mean IBI scores vary (Figure 2-12 and 2-13).

Additional data sets from MBSS 1994 and 1993 sampling were evaluated separately for comparison with the variability analyses presented above. Variability estimates of FIBI scores for replicates within reaches in the 1994 MBSS demonstration study are consistent with results from MBSS 1995-1997. For the 1994 FIBI data, the mean standard deviation and *cv* within reaches across all stream orders are 0.36 and 0.14 respectively, similar to the values of 0.37 and 0.12 for 1995-1997 within-reach variability in FIBI. The average standard deviation and *cv* decreases with stream order (Figures 2-14 and 2-15). As with the 1995-1997 data, variability was slightly higher among sites on first-order reaches. Also, the *cv* was the most stable measure of variability within reaches for the 1994 survey (Figure 2-16).

Sampling protocols for the MBSS 1993 pilot study (Vølstad et al. 1995) differed somewhat from the 1995-1997 survey; therefore, computation of fish and BIBIs for 1993 sites was not feasible within the scope of this study. However, the 1993 study design included many sites within the

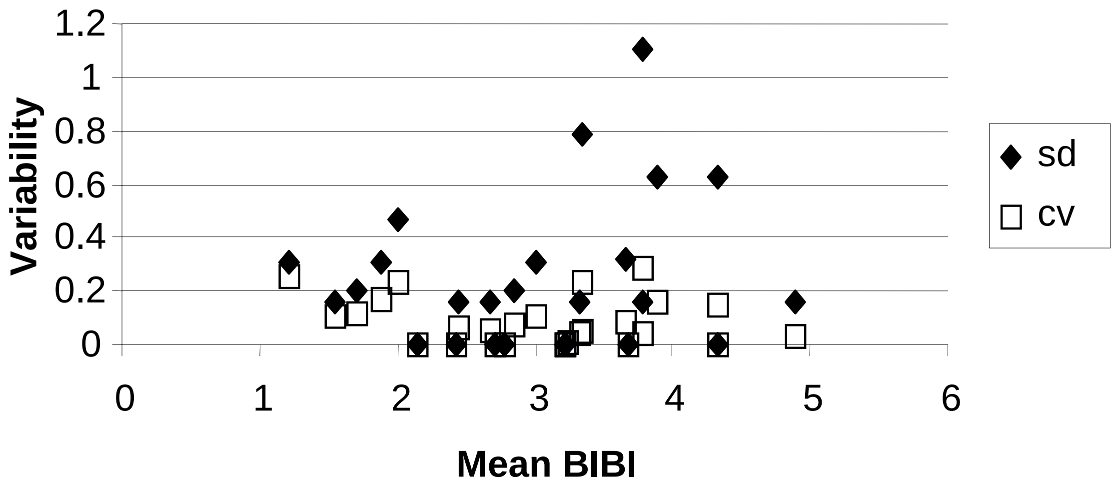


Figure 2-12. Variability (standard deviation and coefficient of variation) between replicate samples within stream segments versus mean BIBI scores, MBSS 1995-1997.

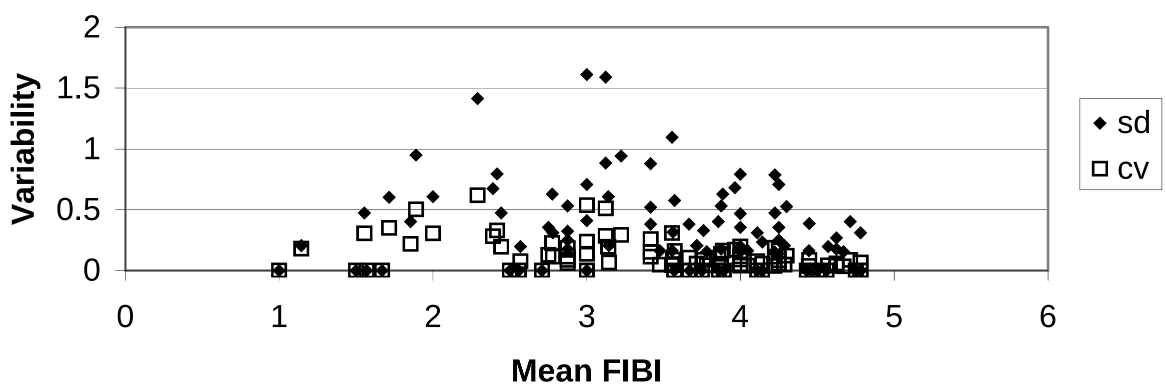


Figure 2-13. Variability (standard deviation and coefficient of variation) in FIBI scores between replicate samples within reaches, MBSS 1995-1997.

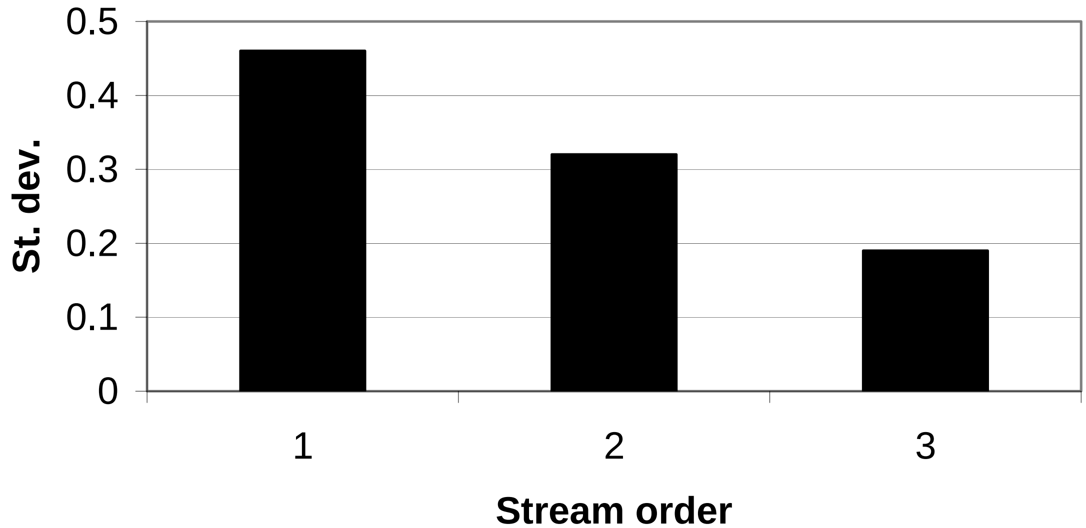


Figure 2-14. Mean standard deviation of FIBI for replicate samples within reaches for MBSS 1994 Demonstration Study. A total of 60 reaches had replicate samples.

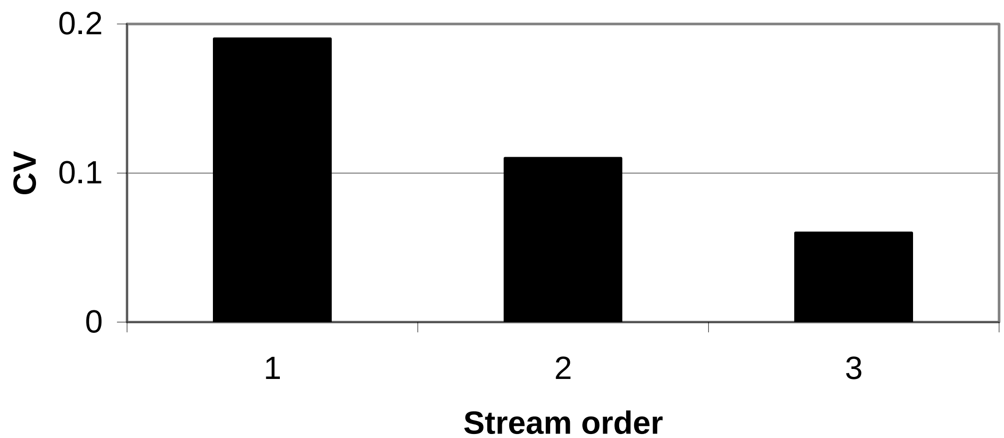


Figure 2-15. Mean *cv* of FIBI for replicate samples within reaches, for MBSS 1994 Demonstration Study. A total of 60 reaches had replicate samples.

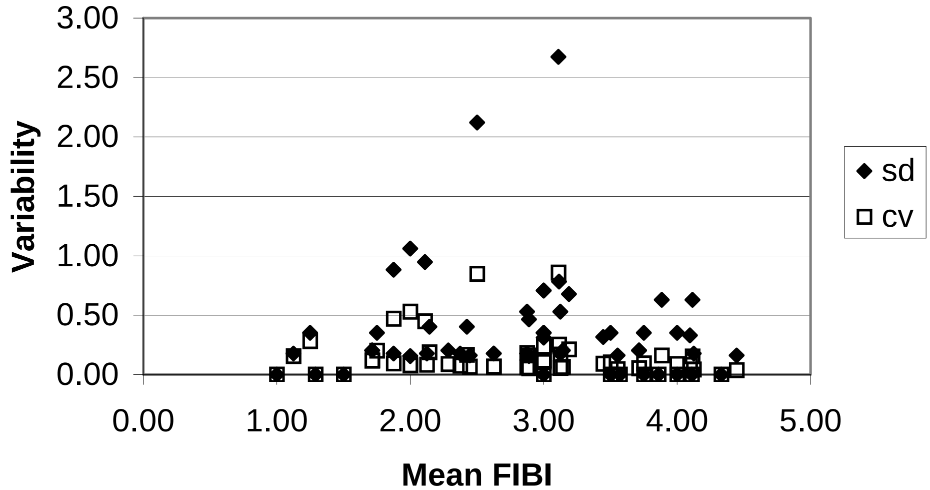


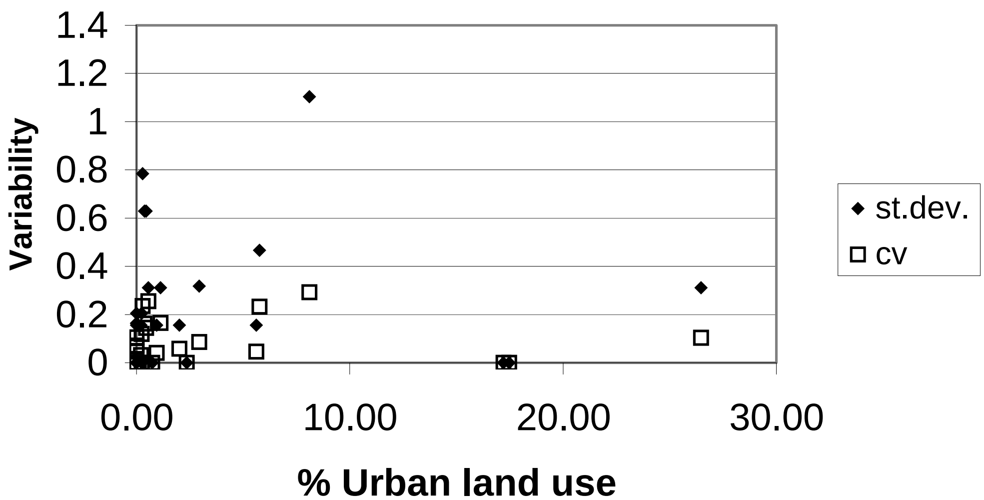
Figure 2-16. Variability versus mean FIBI score, for replicate samples within reaches, MBSS 1994. A total of 60 reaches had replicate samples.

same reaches and was thought to be useful for evaluating local-scale variability in benthic assemblages. A family-level Hilsenhoff biotic index was available for analysis. This index is a weighted average of the pollution tolerance of benthic organisms, ranging from 0 (less tolerant) to 5 (more tolerant). Variability estimates for the Hilsenhoff index, based on replicate samples within reaches for the MBSS 1993, are in Table 2-7. The degree of variability of the Hilsenhoff index within reaches (average cv of 9%) is similar to the variability observed for replicate BIBI scores within segments in MBSS 1995-1997.

Table 2-7. Means of the Hilsenhoff index ($\bar{x}$), standard deviation ($\bar{s}$) and coefficient of variation ($\overline{cv}$) within reaches by stream order using benthic data from MBSS 1993.

		Stream order			
Metric	Statistic	1	2	3	All
	n	24	16	4	44
Hilsenhoff	$\bar{x}$	2.13	2.07	2.00	2.10
	$\bar{s}$	0.11	0.15	0.16	0.13
	$\overline{cv}$	0.07	0.10	0.10	0.09

Based on analysis of the 27 sites with replicate sampling (MBSS 1995-1997), we found no significant relationship between variability in BIBI scores and land use characteristics of the catchment area (Figure 2-17). Hence, the average cv of BIBI scores for single sites (8%) is applied across different land uses. If future data suggest a relationship between land use and variability in IBI scores, mean cv's by land use could be calculated.



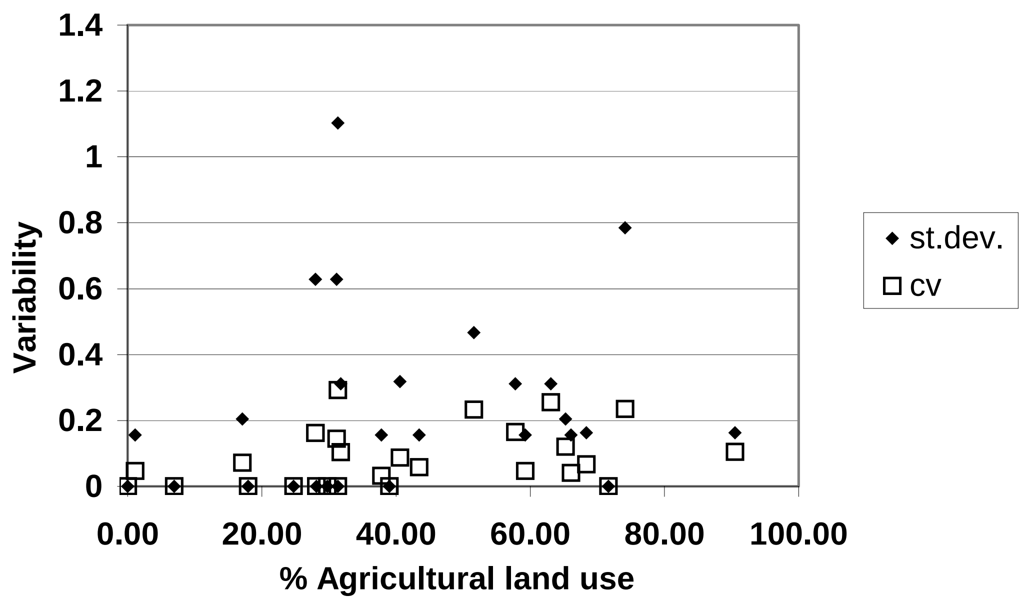


Figure 2-17. Variability in BIBI scores within stream segment different land uses.

2.4.3 Provisional Classification of 12-digit subwatersheds

In MBSS 1995-1997, a total of 451 subwatersheds (12-digit) had one or more sites with IBI scores. Using an average cv of 8% for fish and BIBI scores, 90% one-sided confidence intervals, and the criteria outlined in section 2.3.3 results in the following classification of 12-digit subwatersheds. Of these 451 subwatersheds,

> 287 subwatersheds would be labeled as impaired (93 of these had sites where both fish and BIBI scores failed),
>
> 83 would be candidates for future sampling (i.e., they did not pass or fail), and
>
> 81 would pass.

This simplified example is included for the purpose of illustrating the application of single-site variability estimates to the rating of 12-digit subwatersheds. This example does not account for the assessment of the larger 8-digit watershed, which would also be considered in actual application of the proposed biocriteria framework for Maryland.

3. INTEGRATION OF MBSS AND COUNTY STREAM MONITORING PROGRAMS

3.1 BACKGROUND

In Maryland, both state and local program managers recognize the advantages of integrating stream monitoring. Potential advantages of monitoring program integration include consistent statements to the public about stream condition, increased accuracy in estimates of stream condition, and reduced cost of sampling programs. As programs address increasing needs for information about stream conditions, funding constraints often limit the number of sites that can be monitored. The sharing of data among monitoring programs has the potential to increase the amount of information available to each program.

The partnership between Maryland DNR, U.S. EPA, Montgomery County, and other participants in this integration effort is consistent with the purpose and goals of the Maryland Water Monitoring Council (MWMC). The MWMC was created in 1995 to foster cooperation among the many agencies and organizations responsible for aquatic monitoring across the state. Coordinated approaches to monitoring methods, data management, environmental indicators, and watershed monitoring strategies are encouraged and promoted by the Council. Efforts to integrate the MBSS and Montgomery County stream monitoring programs, as reported below, address specific technical issues related to these general goals.

While the integration of monitoring programs has many benefits, it is important to ensure that the different objectives of local and state programs are met. Successfully integrating those programs requires resolving the following issues: (1) differences in survey design, (2) field sampling protocols and QA/QC, (3) differences in types of data collected, (4) differences in ratings of stream condition, and (5) complexity and cost of data analysis needed to integrate results.

This chapter addresses these issues and other key considerations in the integration of state and county stream monitoring. General guidelines are presented and the integration of the MBSS and Montgomery County, Maryland, stream monitoring programs is used to illustrate the approach. Throughout, the discussion of issues is supported by analysis of existing monitoring data from MBSS and Montgomery County. Further studies required to complete integration and data sharing are proposed, and the Seneca Creek watershed in Montgomery County is recommended as the site of a pilot study to be conducted in 2001. Our intention in this report and in the pilot study is to explore issues relevant both in Maryland and elsewhere in the nation. We hope the lessons learned will serve as examples for other stream monitoring programs. Although Montgomery County is the focus of this study, other jurisdictions (e.g., Prince George's County, Howard County) and organizations (e.g., Maryland Save Our Streams) have already begun to coordinate monitoring efforts with the MBSS.

Since 1994, the Montgomery County Department of Environmental Protection (DEP) has conducted a stream monitoring program to assess the integrity of streams and rivers throughout the County. Since the program's inception, Montgomery County DEP has solicited input from Maryland DNR and other agencies, who serve as members of the County's Biological Monitoring Work Group. The County adopted field protocols and methods recommended by DNR and U.S. EPA at the time of the program's inception. In addition, Montgomery County DEP coordinates with Maryland-National Capital Parks and Planning Commission (M-NCPPC) on site selection and shares information about sites within the County's parklands.

General guidelines for program integration are discussed below, supported with specific examples from the MBSS/Montgomery County integration effort currently in progress. Key steps in the effective integration of multiple programs include the following:

- Review individual program objectives and define goals for integration
- Identify and compile data needed for integration analysis
- Compare sampling frames
- Compare survey designs
- Compare field and laboratory protocols for data collection
- Compare QA/QC protocols
- Compare biological indices (IBIs)
- Develop integrated approach to estimating stream condition

Each of these steps are discussed in the sections below and an outline is proposed for a Seneca Creek pilot study, which if funding can be secured would be conducted jointly by MBSS and Montgomery County in 2001.

3.2 REVIEW INDIVIDUAL PROGRAM OBJECTIVES AND DEFINE GOALS FOR INTEGRATION

There are two key components to this step. First, one must **define the individual and common objectives of the stream monitoring programs being considered**. Some important differences may emerge and it is critical to ensure that each program's objectives are supported by the integration that is developed. The objectives of both programs should be clearly outlined and understood by both parties. Then, **specific goals for integrating** the programs should be defined. The emphasis of the integration may vary depending, for example, on whether goals include (1) reduction of uncertainty via increased sample size or (2) maintenance of existing sample density but reduction of overlapping site locations.

For example, primary objectives of the MBSS include the following:

Assess the current status of biological resources in the state's non-tidal streams (includes derivation of estimates with quantifiable confidence for state, basin, watershed, county, or other subpopulations; examples include mean values, percentages of stream miles exhibiting characteristics of interest, fish population estimates);

Provide biological assessment data to support the development and application of biological criteria (requiring IBI data by 8-digit watershed and within 12-digit watersheds to determine biological impairment; also requires ability to quantify IBI variability);

Quantify the extent to which acidic deposition may be affecting biological resources;

Examine which other water chemistry, physical habitat, and land use factors are important in explaining the current status of biological resources (also useful in biocriteria applications by helping to identify stressors associated with biological impairment);

Compile a statewide inventory of stream biota;

Establish a benchmark for long-term monitoring of trends in stream conditions;

Target future local-scale assessments and mitigation measures needed to restore degraded streams; and

Identify high-quality streams that should be given priority for conservation.

In comparison, primary objectives of Montgomery County's stream monitoring program include the following:

Characterize stream and watershed conditions at finer (subwatershed and areas of homogeneous land use) spatial scales;

Implement long-term monitoring under the requirements of the County's NPDES stormwater permit to monitor the biological, physical, and chemical integrity of County waters;

Assess cumulative impacts to streams at specific locations (targeted reaches or targeted sites);

Assess the impacts of specific developments on the ecological integrity of the County's waters within Special Protection Areas;

Target mitigation measures needed to restore degraded streams;

Evaluate the effectiveness of ecological restoration; and

Identify high-quality streams that should be given priority for conservation.

Note that while many of the goals are similar, one key difference is scale. While one of the main goals of the MBSS is to estimate the status of streams, both statewide and at the 8-digit watershed level, Montgomery County is interested in estimating stream status at much finer scales and in monitoring conditions downstream of particular developed areas.

Several goals were identified for integrating the MBSS and Montgomery County programs in joint discussions. Managers from both programs listed the following key goals for program integration:

Developing consistent statements to the public about stream conditions within Montgomery County;

Increasing accuracy in estimates of stream condition; and

Evaluating whether the two programs duplicate effort and determining the potential for reducing sampling costs.

At a minimum, the integration study should facilitate coordination on stream sampling locations; ideally it will produce a complete integration. There was general agreement that there needed to be a solid technical basis for partnership and that the integrative approach needed to acknowledge and support the individual goals of the two programs.

3.3 IDENTIFY AND COMPILE DATA NEEDED FOR INTEGRATION

It is important to **identify the data needs for integration of stream monitoring programs**. Even to assess integration potential, it is necessary to obtain detailed information (e.g., geographic information system or GIS files, field protocols, raw data) from both programs for careful scrutiny and comparison. If there are existing data from both programs, they can be used to assess potential gains from integrated analyses and coordinated sampling effort. In particular, biological indicator results can be compared (see section 3.6) and joint analytical approaches can be developed and tested (see section 3.7).

Compiling the necessary data is itself not a trivial exercise. Data needs include

> GIS files of streams, watershed boundaries, and all geographic strata used in site selection and indicator development (e.g., ecoregions, subwatersheds, soil types or other regional strata);
>
> Field sampling manual and field data sheets;
>
> IBI scores and procedures for calculating IBIs (if different); and
>
> Complete data needed to calculate county and state IBIs (e.g., raw fish and benthic data, site locations, tolerance and trophic ratings, catchment areas).

Effective integration depends on data consistency among programs. Some data inconsistencies are inevitable, given that programs evolve separately and decisions are made along the way to tailor data collection to fit specific program needs. Consistency issues may be simple (e.g., use of different units for the same parameter) or more difficult (e.g., programs do not collect data on the same parameters) to resolve. Sufficient time should be built into the integration process to resolve these inconsistencies. During integration analyses, good coordination between the data managers from the different programs will help to identify and address discrepancies.

Solid data management and QA/QC practices by both programs will cut down on the difficulties that may be encountered when integrating data. Missing data points or other data errors can complicate the integration of data, increase costs of integration, and cause delays. We recommend that programs adopt rigorous field training and testing, data review, and data entry practices, including double-entry and cross-checks of all field data. For more details on recommended QA/QC procedures, see Kazyak (2000). From the program’s inception, Montgomery County has adopted many elements of the MBSS QA/QC procedures.

Programs under development have an opportunity to coordinate with existing programs by adopting consistent field methods and data base systems, thereby reducing program development costs and technical difficulties in integrating monitoring results. For example, in Maryland, Maryland DNR has published field sampling protocols (Kazyak 2000) and a user-friendly guide to MBSS data (Mercurio et al. 1999) that serve as useful references for a developing local program. In the future, the MBSS will make available its Access data entry programs to county staff. If counties were to adopt the same data entry procedures, significant gains in data consistency would be realized, with substantial cost savings to the counties.

In recent years, advances in software technology have simplified conversions among software programs, so that data file formats can be adapted to programs’ own needs, yet be shared among programs as needed for analysis. For example, data entry could be done in Access, with files subsequently exported to Excel spreadsheets for some uses or to a statistical software package such as SAS for more complex analysis. Note that most of MBSS analysis is conducted in SAS because of its ability to handle multiple types of analyses with large data sets and complex study designs.

3.4 COMPARE SAMPLE FRAMES

The selection of sampling sites in stream monitoring programs is usually based on a particular map of streams in the study area. This base map defines the population of streams to be sampled, otherwise known as the sample frame. For example, the MBSS 2000-2004 sample frame is made up of all first- through fourth-order streams in Maryland, as depicted on the USGS 1:100,000-scale base map, stratified by Maryland 8-digit watershed boundaries. As the stream network and study area boundaries may be somewhat different on different maps, **comparing the sample frames used by different programs** is a critical step in program integration.

The first step is to visually compare stream networks. Two types of differences can occur: (1) individual stream reaches are present on one map, but absent on the other, and (2) the same stream reaches are present, but are more meandering on one map than the other, resulting in greater total stream length. If substantial differences are apparent, a quantitative comparison should be done. By visual inspection, all stream reaches that appear in both sample frames should be identified to create a map layer of this “overlapping” portion of the stream network. The streams unique to each separate map would form separate map layers. These three layers would then be used to compare the numbers of stream miles in each sample frame and in the overlapping streams; these data will be needed to support areawide estimates (means or percent stream miles).

If the survey design for either program uses stream order to stratify site selection, stream orders need to be identified. Stream order refers to a systematic process for describing the degree of branching of a stream network within a watershed (Strahler 1957). The order of any stream segment is determined by starting at the headwaters and labeling each unbranched tributary as order one. Where two first-order streams come together, a second-order stream is created. Similarly, when two second-order streams merge, a third-order stream is created. The junction of any two streams of equal order results in a stream of the next higher order. Stream branching patterns are determined by many factors including geology, soils, relief, precipitation, and the degree to which streams are channelized and piped underground. Determining stream order is a function of map scale and the delineating process used. Different depictions and, therefore, stream orders, will be derived when different scale maps are used. Similarly, if one program uses only the "blue lines" on the quads and another program extends the stream network based on contour crenulations, very different stream orders will result. A universally accepted procedure for delineating tributaries to determinate stream order does not exist (McCammon 1994). For program integration, it is important to define the delineation process. Stream orders, if not already designated on the base maps, can be assigned to each stream reach in the GIS by visual inspection. Quantitative comparisons of the number of stream miles by stream order for each sample frame may be needed to support areawide estimates (means or percent stream miles).

In the case of MBSS-Montgomery County integration, MBSS 2000-2004 sampling uses the USGS 1:100,000-scale base map, while Montgomery County uses a more detailed 1:24,000-scale map. MBSS sampling is conducted within primary sampling units (PSUs) equal to Maryland 8-digit watersheds (or, in some cases, aggregations of two or more small 8-digit watersheds). Montgomery County watersheds are smaller, but are nested within MBSS PSU watersheds. A map of watershed boundaries over the entire county (Figure 3-1) illustrates these differences. Montgomery County watersheds used in the County's baseline monitoring program (e.g., Dry Seneca, Little Seneca) are subunits of MBSS 8-digit watersheds (e.g., Seneca Creek). Note that the County also employs even smaller units (subwatersheds with homogeneous land use) for some assessment and planning purposes; these are typically smaller than Maryland 12-digit subwatersheds.

Some MBSS PSU watersheds are contained entirely within Montgomery County, while others cross County boundaries. Initial analysis for program integration will focus on watersheds entirely within the County, where sampling and analytical concerns will be simpler to address. Because Seneca Creek watershed was selected as the site of a 2001 pilot study for program integration, detailed sample frame comparisons were conducted within this pilot watershed. Future analyses using post-stratification would need to be developed to handle the more complex cases where watersheds cross county boundaries.

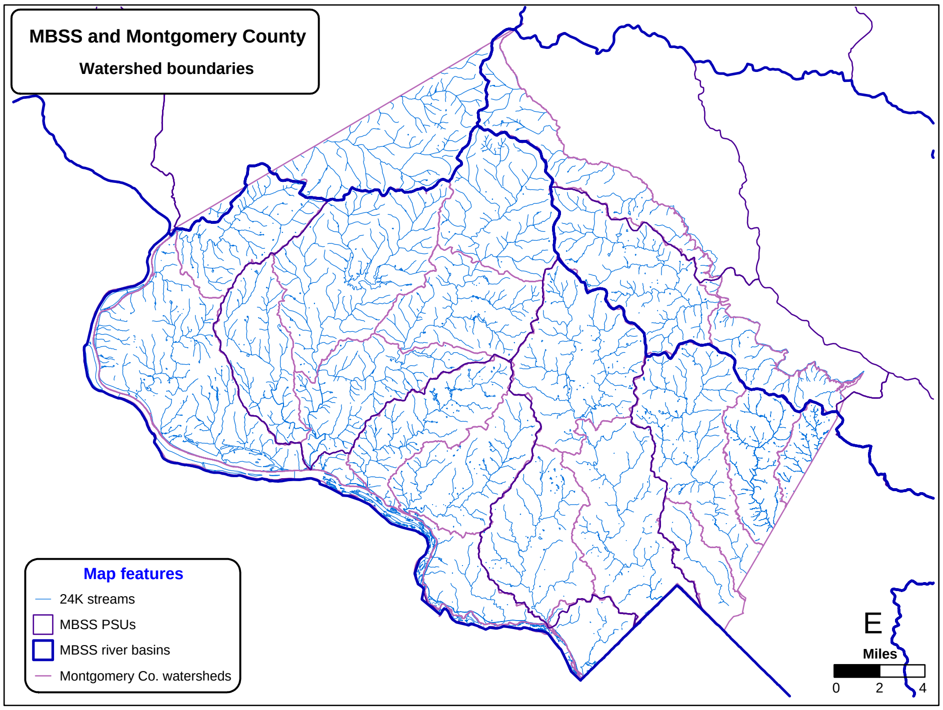


Figure 3-1. Watershed boundaries used in the MBSS and Montgomery County stream monitoring programs. MBSS 2000-2004 sampling is based on primary sampling units (PSUs), which in most cases represent Maryland 8-digit watersheds.

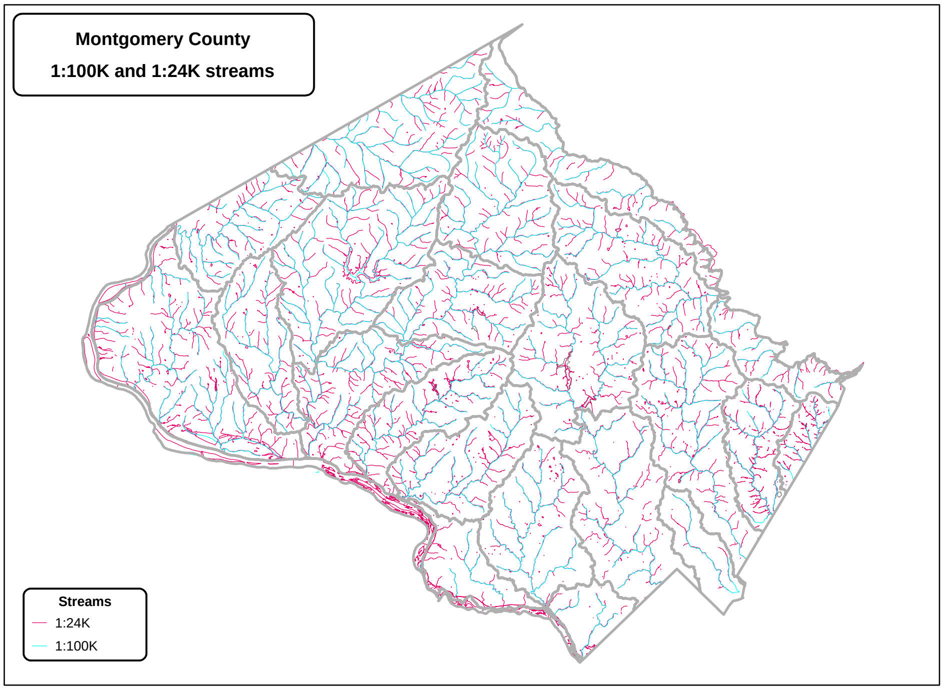


Figure 3-2. Overlay of MBSS 1:100,000 scale and Montgomery County 1:24,000-scale stream base maps.

Using GIS, sample frames were compared first qualitatively by overlaying countywide maps of the two stream networks (Figure 3-2). This visual comparison indicated that substantial differences exist between the MBSS 1:100,000 and Montgomery County 1:24,000 stream base maps. A similar overlay of stream networks within Seneca Creek (Figure 3-3) illustrates differences at a finer level. Much overlap exists, but a greater number of small headwater streams appear in the 1:24,000 map. A small number of streams appear on the 1:100,000 but not the 1:24,000 map. Within Seneca Creek, stream-order designations for some reaches differ, depending on map scale (Figures 3-4 and 3-5).

GIS analysis of streams within the Seneca Creek watershed was used to assess differences between the sample frames. Stream maps were examined for differences for each stream reach (see schematic diagram, Figure 3-6). Overlaps and stream reaches unique to each map were identified (Figure 3-7). The sample frame comparison can be depicted in a Venn diagram (Figure 3-8), which shows a large amount of overlap between the two sample frames (202 stream miles, according to the 1:24,000 map), a fairly substantial number of stream miles found only on the 1:24,000 map (120 stream miles), and a few streams found only on the 1:100,000 map (7 stream miles). Within the overlapping streams, the small difference in total stream length (about 7%) is attributable to a greater degree of meandering represented on the 1:24,000 map scale. Stream lengths by stream order were computed for each map, reaches unique to each map, and for the overlapping area (Table 3-1). In Seneca Creek, total stream length and length by stream order differ substantially between sample frames. Note that both maps are only approximations of the real stream network (the resource of concern); managers of both programs have noted that field crews have detected inaccuracies using both maps. For smaller streams, the maps represent a snapshot in time and may include streams that do not exist at the time of sampling. For example, some mapped streams are actually dry when visited by field crews.

The 1:24,000 map can improve stream coverage and thus provides additional information to characterize all streams, particularly in the subset of small headwaters not appearing on the 1:100,000 map. Where the two sample frames overlap data from both programs can be combined to characterize this portion of streams. The additional Montgomery County data from streams that are represented only on the 1:24,000 map can then be added. Together these components would improve the spatial coverage and hence the accuracy of estimates of stream condition. It is likely that both types of estimates would be useful. In comparing Seneca Creek to the other watersheds in the state (one of the major goals of the MBSS), only the overlapping portion of the sample frame would be used for consistency. In contrast, the best characterization of Seneca Creek itself would encompass data from both the overlap and the 1:24,000-only streams.

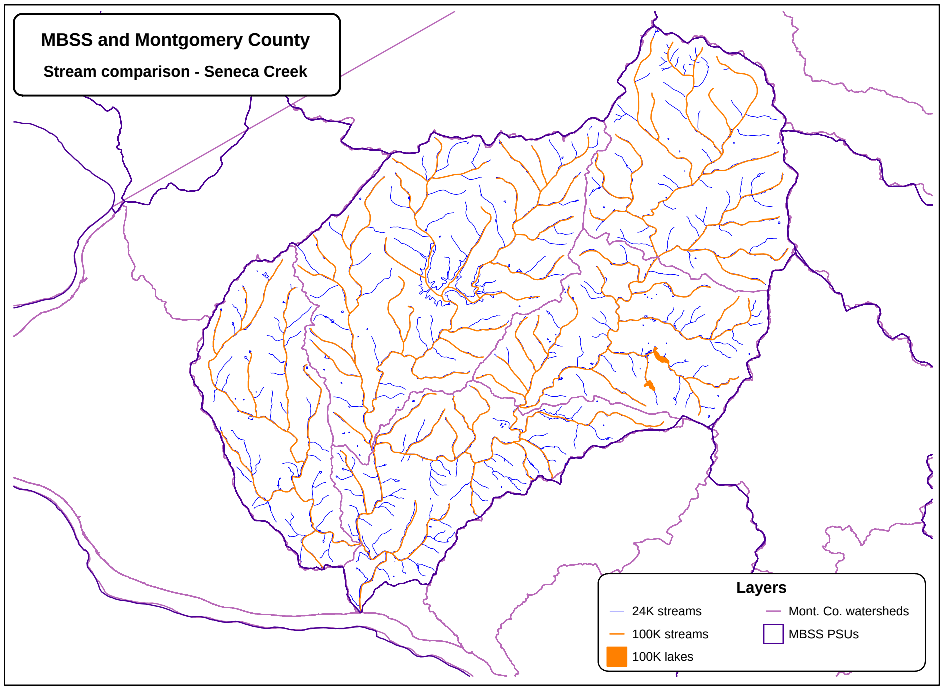


Figure 3-3. Overlay of MBSS 1:100,000-scale and Montgomery County 1:24,000-scale stream base maps within Seneca Creek watershed.

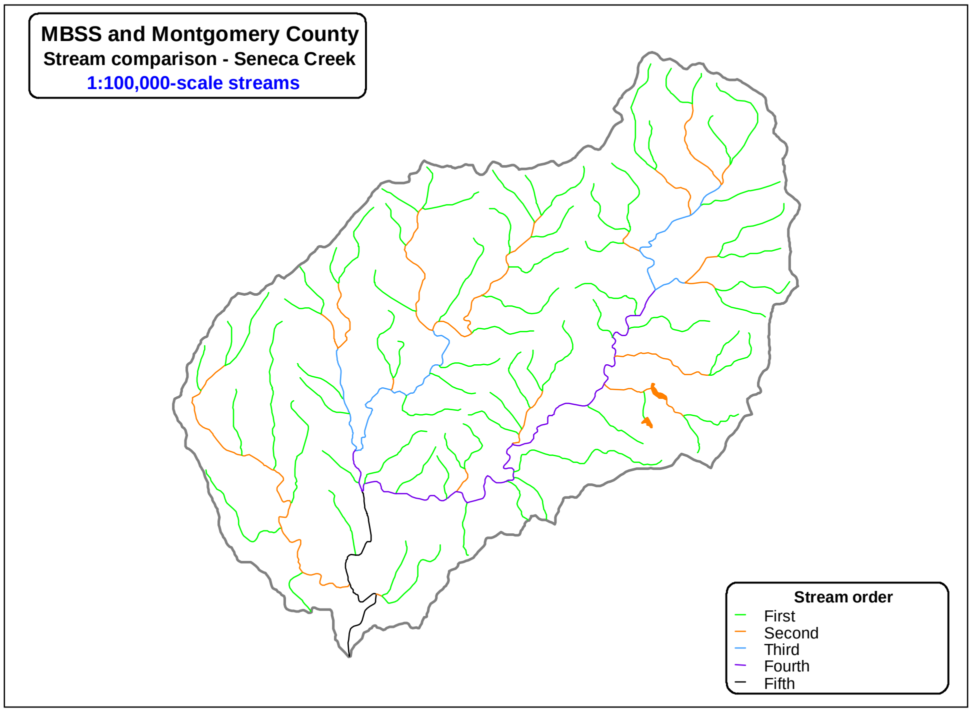


Figure 3-4. MBSS 1:100,000-scale stream base map, with stream order designations.

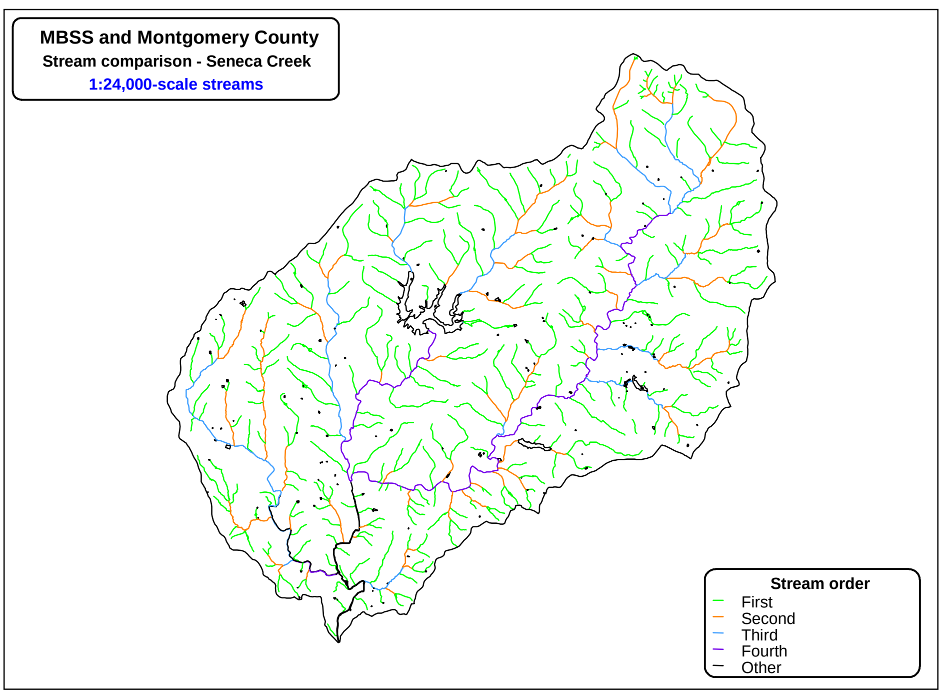


Figure 3-5. Montgomery County 1:24,000-scale stream base map, with stream order designations.

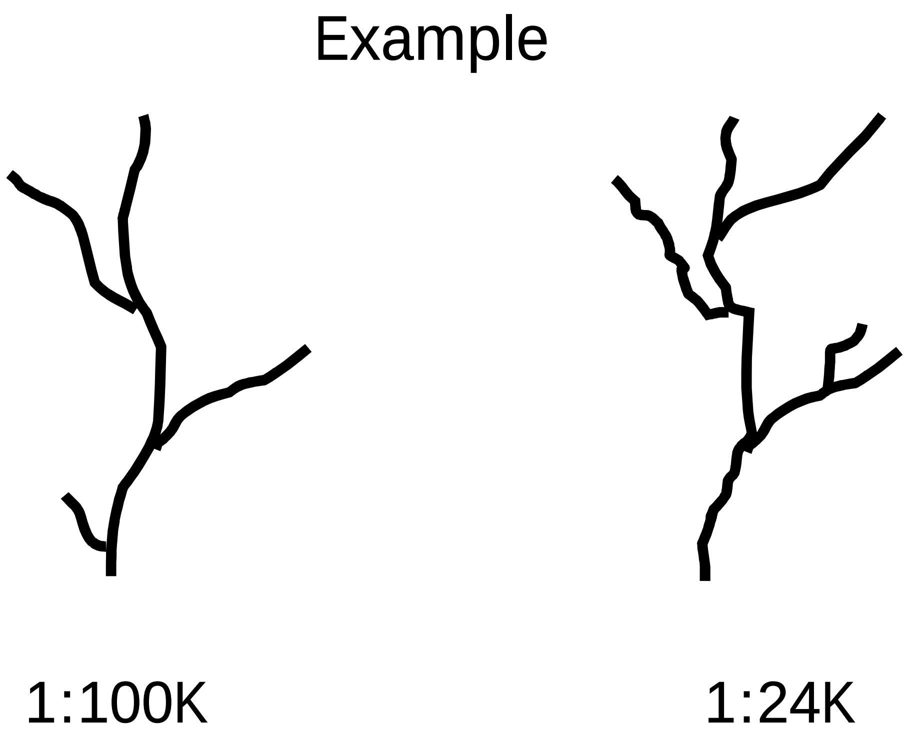


Figure 3-6. Schematic diagram showing hypothetical differences between two sample frames.

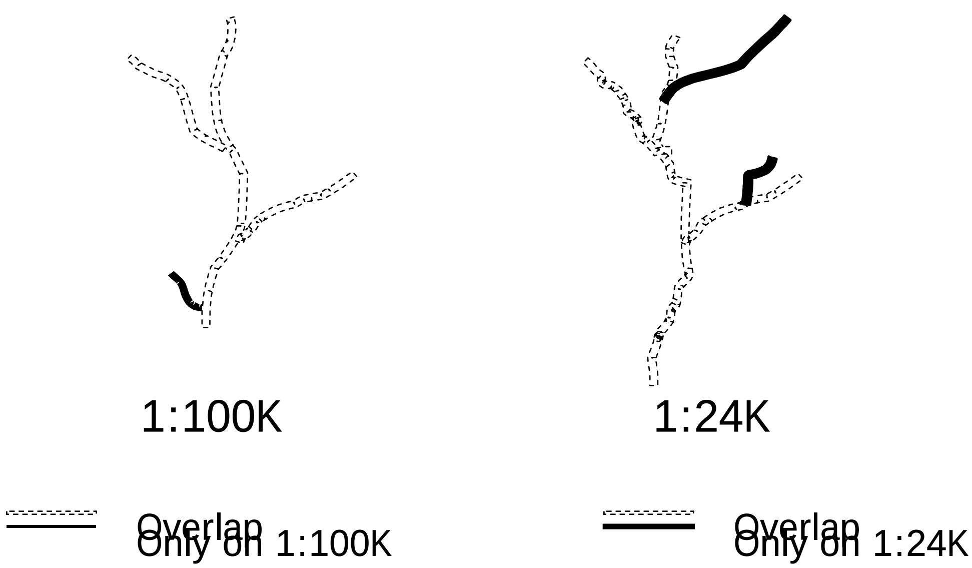


Figure 3-7. Schematic diagram depicting method for identifying stream lengths that overlap on the two map scales and streams unique to each of the two map scales (sample frames).

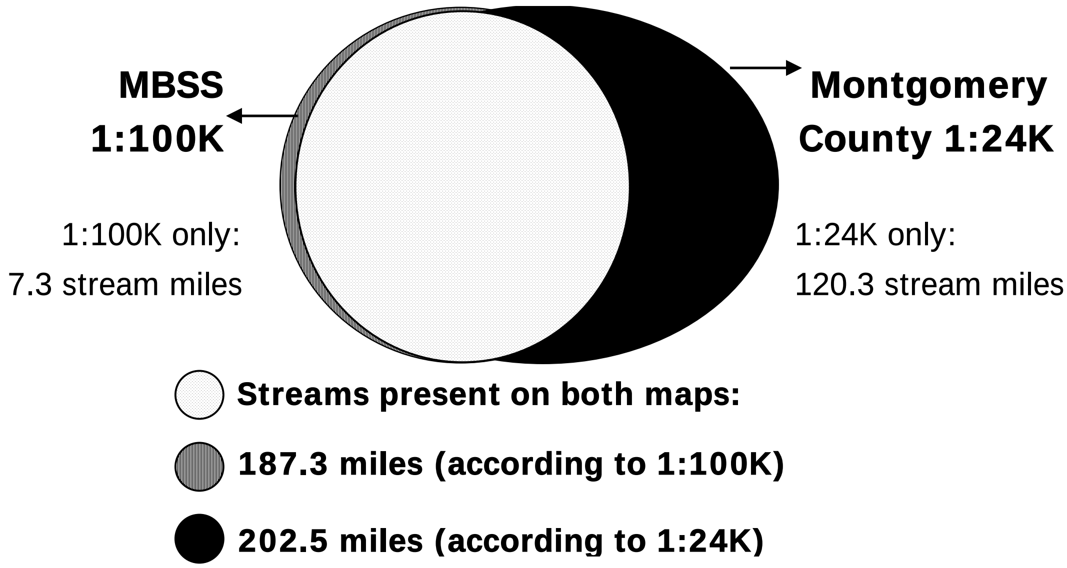


Figure 3-8. Venn diagram illustrating sample frame comparison for Seneca Creek watershed.

Table 3-1. Sample frame comparison: Number of stream miles in Seneca Creek watersheds by stream order. Stream order designations depend on map scale.

Stream Order	In 100K (Total)	In 100K Only	Overlap (according to 24K map)
According to 1:100K Map			
1	126.0	3.0	123.0
2	35.8	4.0	31.8
3	13.0	0.3	12.7
4	14.1	0.0	14.1
5	5.7	0.0	5.7
Total	194.5	7.3	187.3
Stream Order	**In 24K (Total)**	**In 24K Only**	**Overlap (according to 100K map)**
According to 1:24K Map			
1	201.1	114.7	84.5
2	55.4	5.6	49.5
3	32.2	0.0	32.2
4	24.4	0.0	24.4
5	11.9	0.0	11.9
Total	324.9	120.3	202.5

(a) Includes all streams depicted on 1:100K base map
(b) Excludes streams that were found to fall within lakes on 1:24K map

3.5 COMPARE SURVEY DESIGNS

Differences in survey designs can cause major difficulties when integrating monitoring programs. Survey designs are often tailored to the particular objectives and goals of a program, employing selection procedures and analyses that support individual program's management needs. Analysis of data from surveys with different designs can be particularly complex.

MBSS uses probability-based sampling to support areawide estimates of stream condition. In the MBSS Round Two for 2000-2004 (Southerland et al. 2000), MBSS sites are selected within primary sampling units (PSUs) equal to Maryland 8-digit watersheds (or, in a few cases, aggregations of two or more small 8-digit watersheds). Lattice sampling is used to schedule sampling of all PSUs statewide over the five-year survey period. The stream reaches are divided into non-overlapping, 75-meter segments; these segments are the elementary sampling units (sites) from which field data are collected. Stream segments in each PSU are selected using either stratified random sampling with proportional allocation (grouped by 1st-2nd or 3rd-4th order) or simple random sampling (Cochran 1977). This allocation ensures that all stream sites in a PSU have the same probability of being selected. The target sample size in each PSU is a minimum of 10 sites; more samples are allocated to larger PSUs on an ad hoc basis. For example, Seneca Creek is allocated 15 sites for sampling in 2001.

Montgomery County's baseline monitoring program visits watersheds on a rotating basis over five years. The basic watershed unit is typically a third-order watershed as defined by the 1:24,000 map; larger watersheds may be divided into subwatersheds.

Montgomery County uses both targeted and probability-based sampling to support different management needs. Sites are selected in one of three ways, using geographic and stream order stratification: (1) reaches are randomly selected and sites are randomly chosen on the reach, (2) reaches are targeted and sites are randomly chosen on the reach, or (3) both reaches and sites are targeted. For the purposes of developing integrated estimates of stream condition, only the probability-based samples (selection methods 1 and 2) will be considered in this study. Targeted sites are useful for other purposes (particularly to diagnose causes of stream degradation at specific local sites), but do not support area estimates with quantifiable precision. Over time, the Montgomery County program is shifting to random selection of reaches and sites, but will still employ some targeted reaches and fixed sites for trends detection.

Montgomery County's random selection of reaches and sites within reaches (selection method 1) is conducted as follows: First, if large, the watershed may be subdivided into several geographic strata (e.g., subwatersheds). Within each of these strata, streams are stratified by stream order. Within a stratum, stream reach is the primary sampling unit (PSU). A sample of one or more reaches is randomly selected in each stream order (first, second, third, fourth). Note that for variance estimates, a minimum of two reaches per stream order at the lowest watershed subdivision is required. Within each selected reach, a 75-m segment is selected at random. The sampling frame for Seneca Creek is illustrated in Figure 3-9.

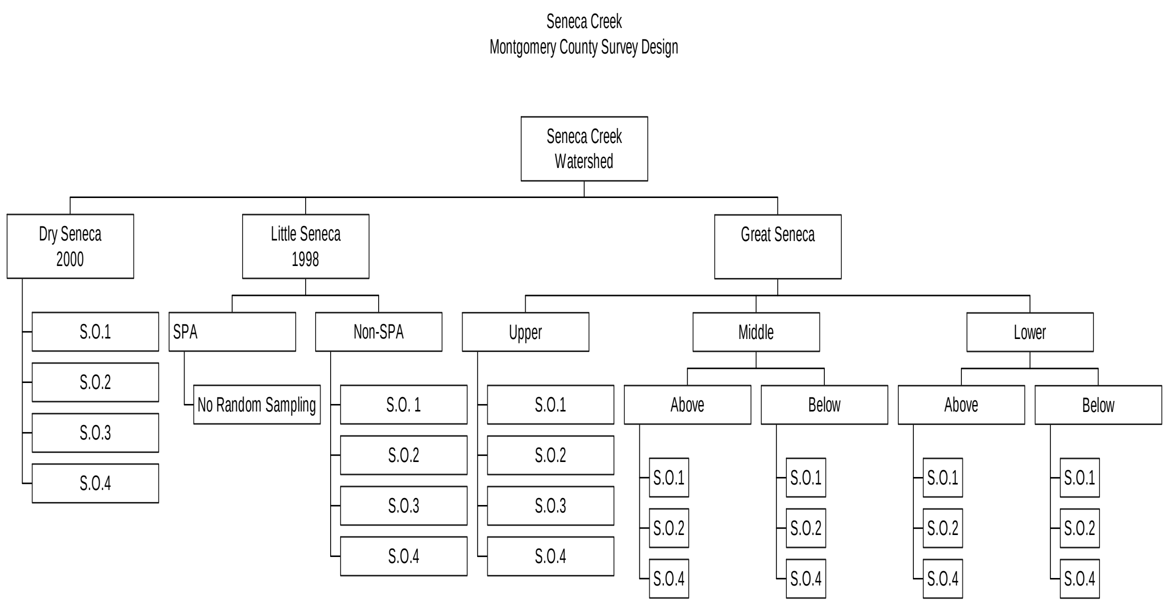


Figure 3-9. Stratification used by Montgomery County to select samples within Seneca Creek watershed. SPA = Special Protection Area. SO = stream order. In Great Seneca, Above and Below refer to lines drawn by Montgomery County staff that best delineate subwatershed boundaries.

Note that the two different site selection approaches each meet the individual goals of their program. MBSS employs probability-based study design that supports statewide, basinwide, and watershed estimates with quantifiable variance (error) estimates. MBSS data are also used for a variety of other purposes, such as research on associations between stream condition and stressors, biodiversity analysis, and identification of impaired waters for 305(b) reporting and 303(d) listing. Montgomery County assesses streams with greater site density, has more need for site coverage of particular areas of interest, and employs fixed stations to detect trends. County data support the Countywide Stream Protection Strategy (a living document) and other County needs for baseline data, restoration targeting, and identification and monitoring of high-quality streams. Stratification within County watersheds assures a good spread of samples (e.g., in upper, mid, lower sections). Random selection of reaches and sites within reaches is a form of probability sampling and supports unbiased estimates of means across reaches. While MBSS generally visits sites once, the County has established fixed sites that will be revisited over time. For example, the County plans to revisit most of its existing Seneca Creek stations sampled in the 1990s in 2001. The fixed sites include stations that initially were randomly selected.

3.6 COMPARE FIELD AND LABORATORY PROTOCOLS FOR DATA COLLECTION

A **detailed comparison of field and laboratory protocols** from both programs should be done. By conducting a detailed side-by-side comparison of protocols, one can evaluate differences and determine what additions to either program are needed to meet joint data needs. To evaluate whether data are recorded in a similar manner, sample field data sheets should be reviewed. It is important that the integration plan maintains the integrity of the data needed to meet the objectives of both programs.

Field manuals and data sheets were reviewed to develop a side-by-side comparison of MBSS and Montgomery County protocols. Results, summarized in Table 3-2, showed that MBSS and Montgomery County programs collect much of the same stream data, but with some important differences.

There are several major differences that may affect IBI scores. Montgomery County did three electrofishing passes during the 1990s, while MBSS does only two. Montgomery County samples benthic organisms with two kick net samples in riffle habitat only, identifies up to 200 benthic organisms in the lab, and only identifies oligochaetes and chironomids to family; while MBSS samples using 20 plots of a D-net in a variety of habitats (primarily riffles), identifies up to 100 organisms in the lab, but mounts and identifies oligochaetes and chironomids to genus or lowest possible taxonomic level.

There are other differences that may be important because of the numerous other uses of the data. MBSS collects more water chemistry data, including laboratory analysis for nutrients and acid-deposition-related measures. The County only collects field measures, but is considering adding to their current list of water quality parameters. The two programs collect similar yet slightly different physical habitat data. Also, the County has established permanent transects to evaluate changes in stream profile over time through detailed geomorphic measures. MBSS collects additional information on amphibians, reptiles, mussels, and aquatic plants; Montgomery County is in the process of developing a new amphibian and reptile monitoring program, but does not currently collect herpetofaunal data.

Table 3-2. Comparison of Montgomery County (Van Ness et al. 1997) and MBSS Round Two (Kazyak 2000) stream sampling protocols.

		Montgomery County	MBSS
Fish	fish sampling	75-m, 3 pass	75-m, 2 pass, number of anodes varies by stream size
	index period for fish	June 1 through mid-Oct (usually do not sample in August if hot or dry weather)	June 1 – Sept 30
	fish biomass	individual trout weighed	total biomass taken for each pass
	game fish data collected	does not specify except for trout (length, weight, fin wear)	length of each game fish taken; game fish are all trout, pike, and bass
Benthos	benthic sampling	2 1-m^2 kick net samples, composited, 200-organism subsample	20 1-ft^2 D-net samples, composited, 100-organism subsample
	benthic habitat sampled	riffles one kick net sample in area of fast current velocity, one in slower	riffle (preferred), rootwad/woody debris/leaf pack, macrophytes, undercut banks (number of plots per habitat type recorded)
	index period for benthos	March15 – April 15 and October 15 – Nov 15	based on degree days; March 1 to approximately May 1
	laboratory identification of benthos	To genus; chironomids/oligochaetes to family	To genus; chironomids/oligochaetes mounted and identified to genus where possible (otherwise to family, subfamily, or tribe)

Table 3-2 (con't)

		Montgomery County	MBSS
Water Chemistry	WQ parameters	temp, DO, pH, conductivity taken in summer (field measures)	pH, ANC, sulfate, nitrite, nitrate, ammonia, total N (dissolved and particulate), orthophosphate, total P (dissolved and particulate), chloride, conductivity, and DOC in spring (lab analysis) DO, pH, conductivity, turbidity in summer (field measures) temperature measured continuously during summer
Habitat	qualitative habitat assessment	similar approach; slight differences in list of parameters and scoring guidelines	
	habitat inventory	detailed habitat inventory included for in-channel features (e.g., number and length of pools/riffles)	checkboxes for stream character features, relative abundance
	cross sections and other physical measures	velocity/depth profile taken at permanently marked cross-section and used to develop discharge/stage relationship wetted width, channel width, thalweg depth, bank height, bank material, % bank height with vegetation, vegetation type, % canopy cover depth, etc. done at three transects (0, 37.5, 75 m)	velocity/depth profile taken at cross-section wetted width, thalweg depth, velocity at four transects (0, 25, 50, 75 m) max depth
	altitude at site		recorded w/ altimeter

Table 3-2 (con't)

		Montgomery County	MBSS
	substrate	Wolman Pebble Count % embedded - left, center, right	pebble count not done single estimate of % embedded from riffle if present
	riparian buffer width	detailed analysis of riparian buffer countywide, using aerial photography and GIS	overall estimate of width on right, left banks buffer type and adjacent land use type and severity of buffer breaks dominant stem size of riparian vegetation estimated
	channelization		evidence of channel straightening or dredging (Y/N) type and linear extent for right and left bank, stream bottom
	local land use		presence of land use types
	gradient / sinuosity		water surface slope between upper and lower end of segment straight line distance
Habitat	bank condition	bank material type	linear extent, severity, areal extent of bank erosion
	bar formation		bar formation (extent and composition)
	exotic plants		relative abundance and species name
	woody debris		number of instream and dewatered woody debris and rootwads
	stream blockages		height and type
	culverts		presence and width
Other Taxa	other taxa		herpetofauna (spring and summer), presence/condition of mussels (summer), relative abundance of aquatic plants (summer)

If field sites are to be shared between programs, coordination needs to ensure that all data required by either program are collected. MBSS and Montgomery County program managers have agreed that for the 2001 field season, complete field data for both programs would be collected. This would allow for some side-by-side comparisons to be conducted through a pilot study to assess comparability (e.g., of different benthic sampling methods). In the future, field effort might be reduced by eliminating some parameters, but only after a clear demonstration that the data collected would serve both programs.

Sharing the field effort could be one way to reduce field costs without sacrificing data completeness. For example, in the future, MBSS and Montgomery County could jointly sample a site as follows:

> Chemistry - share ambient water quality (Hydrolab) data; MBSS collect samples and test for full suite of analytes
>
> Habitat - each program collects own parameters
>
> Fish - complete 2 electrofishing passes - either crew could collect (or employ joint field crew)
>
> Benthos - conduct side-by-side comparison of 2 field methods

3.7 COMPARE AND CALIBRATE BIOLOGICAL INDICES

One of the goals of stream monitoring program integration is to develop consistent assessments of ecological condition, a process made simpler by using the same or consistent indicators of biological integrity. Where possible, programs may agree to use identical indicators or to embark on joint development of indicators. Otherwise, particularly where multiple programs have already developed different indicators and wish to maintain their use to facilitate trends analysis, a comparison and calibration of indicators is needed.

There are multiple factors that contribute to differences in biological indicator results across programs. Field and laboratory protocols are an obvious source of differences. In addition, indicators may differ in metrics selected, metric thresholds, scoring protocols, and the interpretation of index scores. Even when indicator construction differs, the key question is whether they accurately and consistently rate stream condition. If not, one should determine the cause of differences in ratings, so appropriate calibration or adjustments may be made to yield more consistent ratings. When possible, analysis of existing data should be used to best understand the degree and nature of indicator differences.

Maryland DNR and Montgomery County have developed fish and BIBIs that differ in various ways. For detailed information on the IBIs, see Roth et al. (2000), Stribling et al. (1998), and Van Ness et al. (1997). The most recent version of the Montgomery County provisional IBIs (Van Ness, personal communication) were used in the analyses for the report. Metrics for both programs are listed in Table 3-3, along with notes relevant to program integration. If a metric could not be calculated because data were not collected (e.g., it was not possible to calculate the MBSS biomass metric from Montgomery County fish data), the metric was dropped from that analysis.

Table 3-3. MBSS and Montgomery County IBI metrics.

I. BIBIs	
MBSS Non-Coastal Plain IBI	**Notes**
Total number of taxa	Numbers may be lower in Montgomery County due to grouping of chironomids and oligochaetes by family
Total number of EPT taxa	
Total number of ephemeroptera taxa	
Total number of diptera taxa	Numbers will be lower in Montgomery County since chironomids are diptera species
Percent ephemeroptera	
Percent tanytarsini	Cannot be calculated in Montgomery County since chironomids are not further identified
Total number of intolerant taxa	
Percent tolerant taxa	
Percent collectors	
Provisional Montgomery County IBI	
Total number of taxa	
Biotic index	
Ratio of scrapers (scrapers + filtering collectors)	
Proportion of hydropsyche and cheumatopsyche/total EPT individuals	
Proportion of dominant taxa	
Total number of EPT taxa	
Proportion of total EPT individuals	
Proportion of shredders	

Table 3-3 (con't)

II. FIBIs	
MBSS Eastern Piedmont IBI[a] Number of native species – adjusted for watershed area Number of benthic species – adjusted for watershed area Number of intolerant species – adjusted for watershed area Percent tolerant fish Percent abundance of dominant species Percent generalist, omnivores, and invertivores Number of individuals per square meter Biomass per square meter Percent lithophilic spawners	Montgomery County does not record total biomass
MBSS Highlands IBI[b] Number of benthic fish species – adjusted for watershed area Number of intolerant fish species – adjusted for watershed area Percent tolerant fish Percent generalists, omnivores, and invertivores Percent insectivores Percent lithophilic spawners	
Provisional Montgomery County IBI Total number of species Total number of riffle benthic insectivores Total number of minnow species Total number of intolerant species Proportion of tolerant individuals Proportion of omnivores/generalists Proportion of pioneering species Total number of individuals (excluding tolerants) Proportion with disease/anomalies	
(a) applies to Montgomery County watersheds below Great Falls (b) applies to Montgomery County watersheds above Great Falls	

Narrative ratings for the IBIs are as follows:

MBSS IBIs:

IBI 4.0 - 5.0 Good
IBI 3.0 - 3.9 Fair
IBI 2.0 - 2.9 Poor
IBI 1.0 - 1.9 Very Poor

Provisional Montgomery County FIBI:

IBI 4.5- 5.0 Excellent
IBI 3.3 - 4.4 Good
IBI 2.2 - 3.2 Fair
IBI 1.0 - 2.1 Poor

Provisional Montgomery County BIBI:

IBI 36-40 Excellent in Channery Silt Loam ecoregion (35-40 in Silt Loam ecoregion)
IBI 26-35 Good (both ecoregions)
IBI 17-25 Fair (both ecoregions)
IBI 8-16 Poor (both ecoregions)

As an initial step, existing data from MBSS and Montgomery County monitoring programs were used to evaluate differences associated with the two program's different field methods and IBIs.

3.7.1 BIBI Comparability

3.7.1.1 Analysis of MBSS Data

To isolate differences resulting from data analysis and IBI conventions, while controlling for field methods, MBSS benthic data were used to compare the MBSS BIBI and Montgomery County BIBI. BIBIs were calculated using raw data from 63 sites sampled by the MBSS in Montgomery County during the 1995-1997 Survey. We compared IBI scores via scatter plots and linear regression. IBI narrative ratings were used in categorical analysis, based on the narrative rating systems developed by each program (as described above) and assuming the corresponding categories were equivalent in interpretation (e.g., assuming the MBSS rating of good is equivalent to Montgomery County's excellent). Further calibration may be needed to establish appropriate thresholds of equivalency. The two programs should also consider adopting a consistent narrative rating system (use the same category names) to improve clarity in communicating results to the public.

While some sites tended to receive similar ratings, some inconsistencies were apparent (Figure 3-10). Contributing factors may include use of different metrics, use of 100 vs. 200 organism subsamples, and differences in taxonomic level used to identify oligochaetes and chironomids.

We were able to test the effect of more coarse taxonomic identification of oligochaete and chironomids by grouping "lumping" MBSS oligochaete and chironomid data to mimic the Montgomery County protocol. When these taxa were grouped to family level, as specified in the Montgomery County protocol, the MBSS and Montgomery BIBIs were more similar, as expected, but some variability remained (Figure 3-11). Using grouped taxa had a dramatic effect on MBSS BIBI scores (Figure 3-12). With the original BIBI, 10% of sites were rated as good and 44% as fair. In contrast, the BIBI from grouped taxa resulted in no sites rating as good and 25% as fair. A smaller percentage of sites were rated as very poor (22%) by the original IBI than with grouped taxa (44%). The *total number of taxa*, *number of diptera taxa*, and *number of intolerant taxa* metrics are affected by grouping chironomids and oligochaetes to the higher level; percentage metrics could also be affected.

Although field data from these studies do not allow direct comparisons of 100 vs. 200 organism subsamples, the potential effect of subsample size was explored through theoretical analyses (see Section 3.7.1.4).

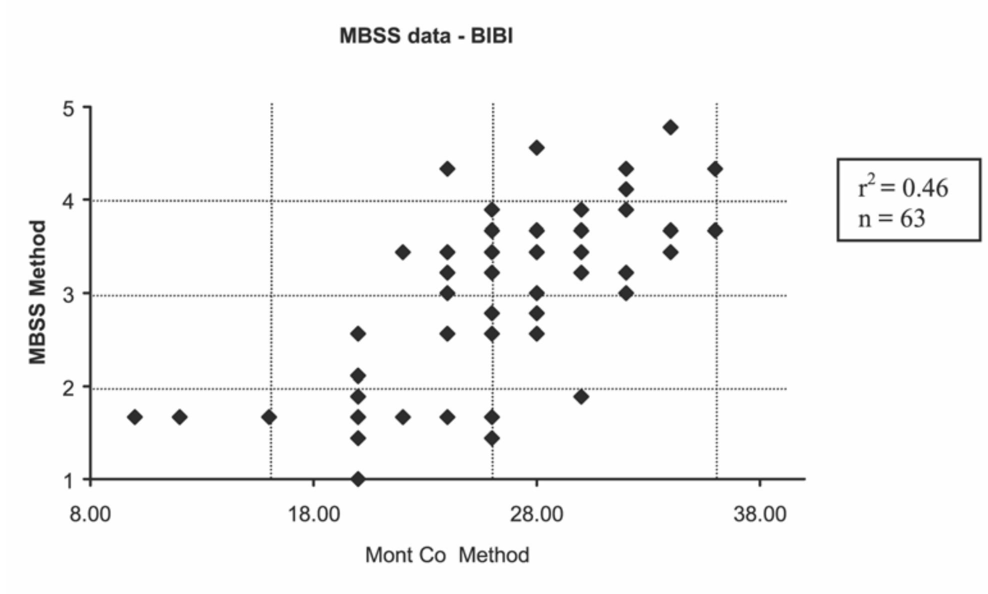
3.7.1.2 Analysis of Montgomery County Data

As an additional test to isolate differences resulting from data analysis and IBI conventions (while controlling for field methods), Montgomery County benthic data were used to compare the MBSS BIBI and Montgomery County BIBI. BIBIs were calculated using raw data from 159 sites sampled by Montgomery County during 1995-1999. We compared IBI scores via scatter plots and linear regression and compared IBI narrative ratings using categorical analysis.

While some sites received similar ratings, some inconsistencies were apparent (Figure 3-13). As above, contributing factors may include use of different metrics, use of 100 vs. 200 organism subsamples, and differences in taxonomic level used to identify oligochaetes and chironomids.

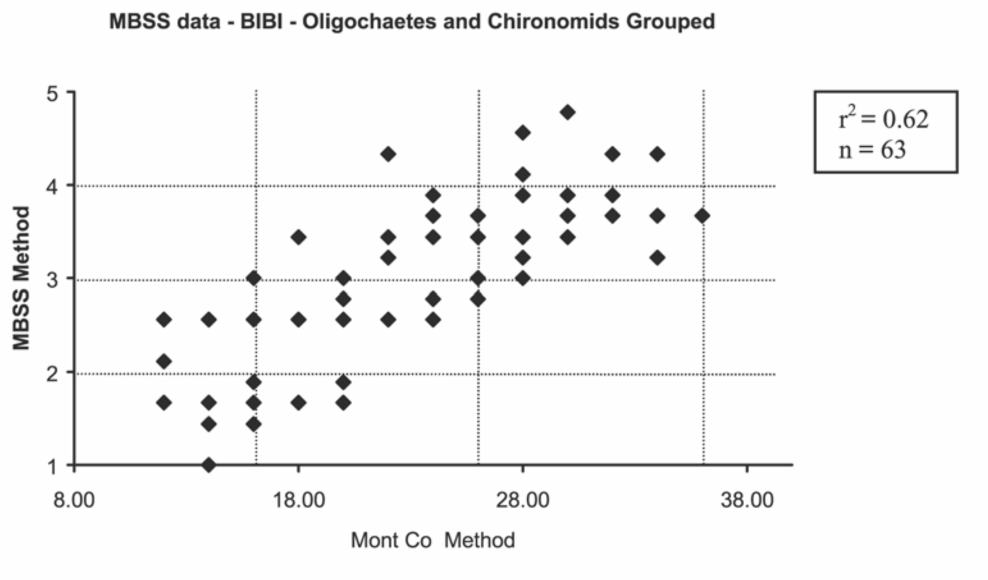
3.7.1.3 Analysis of 1997 MBSS-Montgomery County Joint Field Study Data

In 1997, a joint sampling study was conducted by Montgomery County and MBSS at a small number of sites. Benthic samples were collected using the methods of both programs; fish data were collected jointly. This preliminary integrated field study was conducted to compare monitoring procedures and IBI results. Benthic macroinvertebrates were sampled at twelve locations (ten stream segments where coordinated MBSS/County sampling was done, plus two reaches that by chance contained sites sampled by both programs).



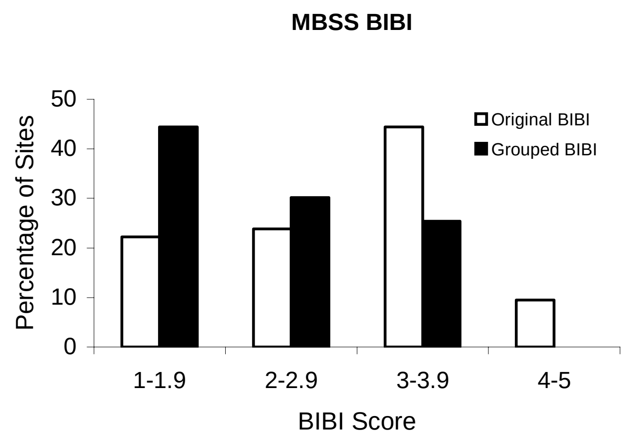
BIBI – MBSS Method					
	Good	0	1	**4**	**1**
	Fair	0	6	**21**	**1**
	Poor	0	**6**	9	0
	Very Poor	**3**	7	4	0
		Poor	Fair	Good	Excellent
	BIBI – Montgomery County Methods				

Figure 3-10. Comparisons of MBSS and Montgomery County BIBI numeric scores and narrative ratings (each program uses different naming conventions for the ratings but they are comparable as aligned in the table), using MBSS data.



BIBI – MBSS Method					
	Good	0	1	**5**	**0**
	Fair	1	9	**17**	**1**
	Poor	4	**9**	2	0
	Very Poor	**10**	4	0	0
		Poor	Fair	Good	Excellent
	BIBI - Montgomery County Methods				

Figure 3-11. Comparisons of MBSS and Montgomery County BIBI numeric scores and narrative ratings (each program uses different naming conventions for the ratings but they are comparable as aligned in the table), using MBSS data, but grouping oligochaetes and chironomids to family level.



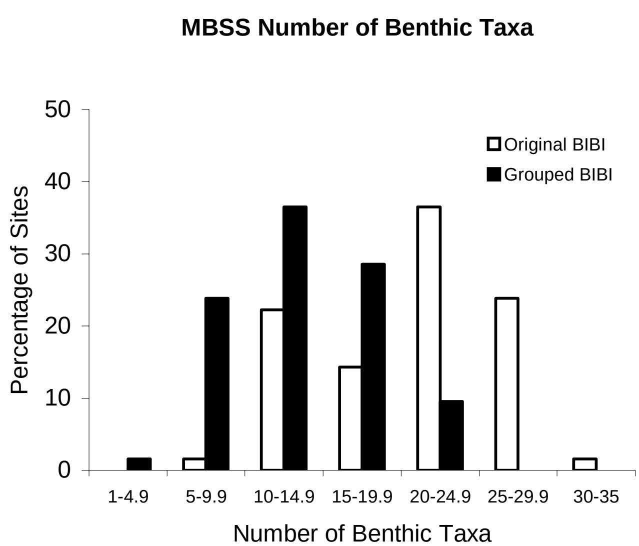
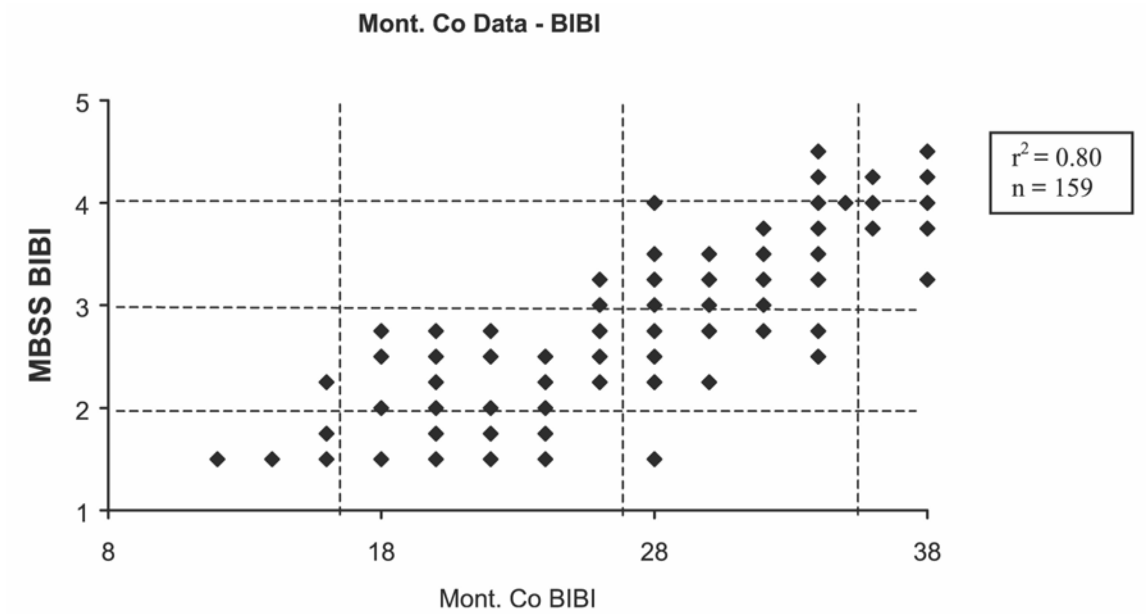


Figure 3-12. Effect of grouping oligochaetes and chironomids on BIBI scores and total number of taxa, using MBSS sites.



BIBI – MBSS Method	Good	0	0	**7**	**26**
	Fair	0	0	**39**	**9**
	Poor	1	**23**	16	0
	Very Poor	**16**	21	1	0
		Poor	Fair	Good	Excellent
	BIBI - Montgomery County Methods				

Figure 3-13. Comparisons of MBSS and Montgomery County BIBI numeric scores and narrative ratings (each program uses different naming conventions for the ratings but they are comparable as aligned in the table), using Montgomery County data.

Results from this joint sampling were used to assess differences in IBI results. Note that differences may result both from the differences discussed above and also, for benthos, differences in field sampling protocols used by the two programs. Montgomery County collects benthic macroinvertebrates within each sampling segment using a 1-m^2 kick net with a mesh size of 530 microns (Van Ness et al. 1997). Riffles are sampled using the kick net to collect benthos from an approximately 2-m^2 composite area. Two samples are collected per stream segment, one from an area of fast current velocity and one from an area of slower current velocity. The two samples are then pooled and a representative subsample of about 200 animals is identified in the laboratory. MBSS collects 20 D-net samples from riffles and other habitats, composites them, and then identifies a subsample of about 100 individuals selected from random grid cells (Kazyak 2000).

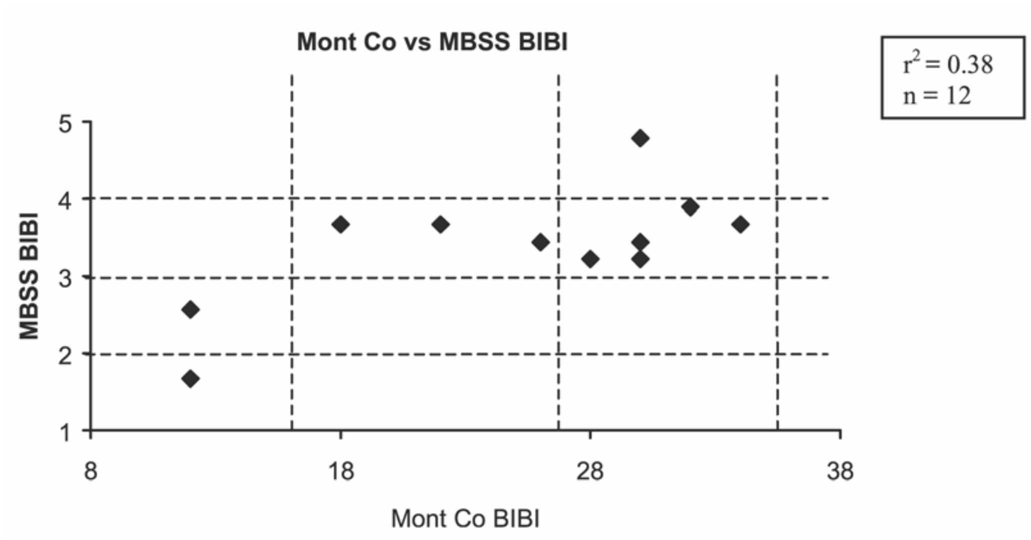
BIBI results from both programs were compared at the sites sampled in the joint study. First, MBSS data were scored using the MBSS BIBI; Montgomery County data were scored using the Montgomery County IBI. This comparison includes natural variability within the stream segment (i.e., benthic macroinvertebrates were collected from different locations within the segment) and sample method variability (kick-net vs. D-net), as well as all the differences that result from the use of two different data and IBI conventions (i.e., subsample size, taxonomic level, metrics used). Linear regression results comparing the two IBIs suggested some similarity, but with high variability. Results were inconclusive, because data were limited to a small number of sites (Figure 3-14). Categorical analysis was also inconclusive, but suggested that further investigations might prove useful in evaluating comparability.

Next, sampling results were reanalyzed to remove the effect of using different IBIs; MBSS data and Montgomery County data were both scored using the Montgomery County BIBI. Oligochaetes and chironomids were grouped prior to scoring. Results were again somewhat encouraging, but inconclusive (Figure 3-15). Note that 4 out of 12 sites received the exact same score; however, differences at other sites resulted in a highly variable relationship, with r^2 similar to that of the previous comparison.

Results of several statistical tests (described in Section 2) showed that differences between MBSS and Montgomery County results from these shared sites were substantially greater than those for MBSS replicate samples (Table 3-4). These preliminary results indicate the need for further field study of sites across a broader range of conditions.

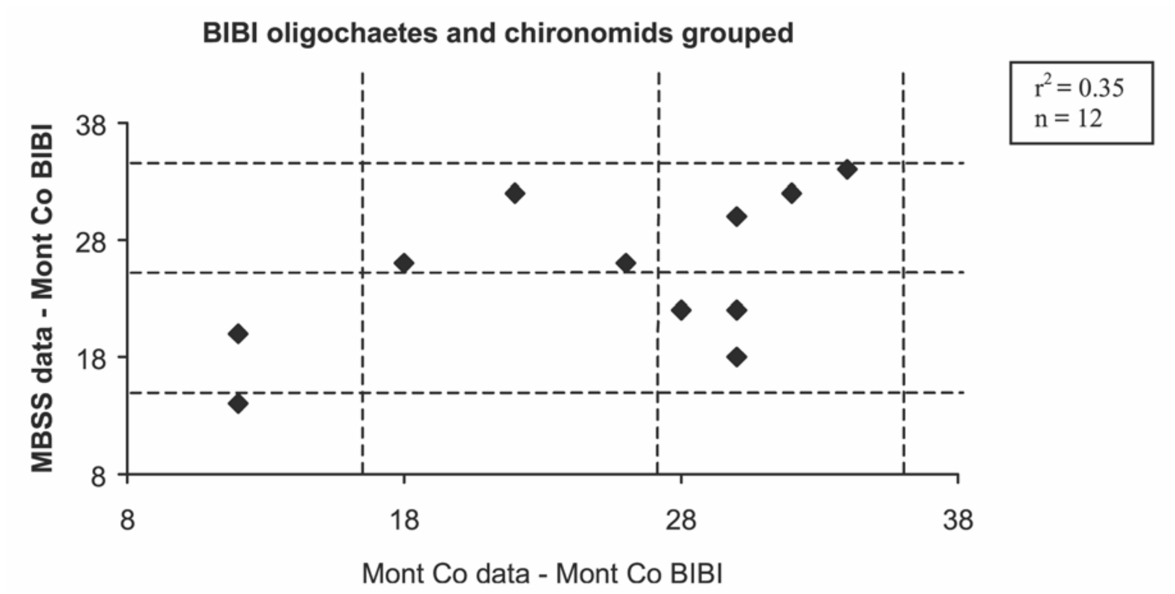
Table 3-4. Measures of reliability of IBI scores for replicate samples within segments.

IBI	DATA	n	R^2	θ	*Cronbach's alpha*	κ_{simple}	κ_{wt}	ρ
Benthic	MBSS 1995-1997	27	0.72	0.85	0.92	0.57 (0.13)	0.70 (0.09)	0.91 (0.06)
Benthic	MBSS-Montgomery County	12	0.37	0.67	0.84	0.21 (0.20)	0.45 (0.19)	0.86 (0.14)
Fish	MBSS-Montgomery County	11	0.34	0.55	0.48	0.06 (0.22)	0.15 (0.20)	0.42 (0.34)



BIBI – MBSS Method					
	Good	0	1	**1**	**0**
	Fair	0	3	**4**	**1**
	Poor	1	**0**	0	0
	Very Poor	**1**	0	0	0
		Poor	Fair	Good	Excellent
	BIBI - Montgomery County Methods				

Figure 3-14. Comparison of separate Montgomery County and MBSS BIBI scores, 1997 joint sampling study (each program uses different naming conventions for the ratings but they are comparable as aligned in the table).



MBSS Data Montgomery Co. BIBI	Excellent	0	0	**0**	**0**
	Good	1	1	**3**	**1**
	Fair	0	**1**	4	0
	Poor	**1**	0	0	0
		Poor	Fair	Good	Excellent
	Montgomery County Data, Montgomery County BIBI				

Figure 3-15. Comparison of MBSS data and Montgomery County numeric data, both assessed with Montgomery County BIBI (oligochaetes and chironomids grouped), 1997 joint sampling study (each program uses different naming conventions for the ratings but they are comparable as aligned in the table).

3.7.1.4 Effects of Subsample Size

Composite samples of freshwater benthos often contain a large number of organisms, and it would be very expensive to sort, identify, and count all organisms in the laboratory. For such reasons, sorting and species identification in the laboratory is limited to a representative subsample of organisms from each composite sample. For a fixed survey cost, there is a trade-off between the number of organisms in the subsample and the number of stream segments that can be sampled by the field crews.

Fixed-count subsampling can provide reliable estimates of taxa richness (e.g., the number of taxa per standard number of organisms). For a given species composition in a composite sample, a subsample based on a constant number of organisms yields consistent estimates of taxa richness (Barbour and Gerritsen 1996). For a fixed number of stream segments, doubled or tripled subsampling effort may result in some improvement in the ability to classify watersheds as degraded or non-degraded because of improved detection of rare species. However, for a fixed survey cost such increased subsampling effort would force a reduction in the number of sites sampled in a watershed. Our analyses (detailed in Section 2) indicate that the variability in IBI scores from replicate samples within stream segments is low relative to IBI scores among stream segments. If this is the general case, the precision in mean IBI scores for a watershed is more determined by the number of sampled stream segments and less by the number of replicate composite samples within stream segments. The large number of plots constituting the composite sample and the fixed count of 100 organisms for establishing IBI scores appears to characterize a stream segment fairly accurately. This suggests that the 100-organism count is sufficient for characterizing Maryland streams. Empirical studies have shown that species richness as a function of sample size reaches an asymptotic level for a count of between 100 and 900 organisms depending on overall richness in the sample (May 1975, Barbour and Gerritsen 1996). Somers et al. (1998) conclude that counts of 100 animals are sufficient to distinguish the littoral benthic communities of small inland lakes in south-central Ontario. Although 100 organisms appears to characterize Maryland streams well, a greater number would increase the precision in taxa richness metrics. However, a larger count would also add significant laboratory costs. We have outlined a comparison study in section 3.8.3 that would address the issue of subsample size.

Fixed-count subsampling is employed in the MBSS, with a target of selecting 100 organisms from each composite sample for identification. Subsamples of a fixed number of organisms yield an estimate of numerical species richness (Barbour and Gerritsen 1996). A sorting pan with grids is employed to achieve a representative sample of organisms from the composite. Organisms from a random selection of “grids” within a pan are sorted from the entire sample. For a well-mixed composite sample, this procedure will, approximately, produce a simple random sample of organisms. Assuming simple random sampling of organisms, the binomial distribution can be used for evaluating the effects of sample size on the probability of including taxa in the subsample.

Let us assume that the total number of organisms N in a composite sample is large relative to the subsample of n organisms. If taxon X constitutes a proportion P of the total number of organisms N, then the probability that a random subsample of size n contains exactly a organisms of X is (Cochran 1977)

$$\Pr(a) = \frac{n!}{a!(n-a)!} P^a (1-P)^{n-a}$$

and, hence, the probability that at least one organism of X is in the subsample is

$$\Pr(a \geq 1) = 1 - \Pr(0) .$$

We used this formula to estimate how the probability of including a taxon relates to its proportion P of the composite and the subsample size n. We also calculated the subsample size n required to detect taxa with 90% probability for varying P of the taxa.

Here we present theoretical results showing the chance of detecting taxa in subsamples of varying size, given occurrence rates (e.g., taxa with a relative abundance that accounts for x% of total number of organisms). The theoretical example assumes simple random sub-sampling of well-mixed composite sample. In practice, if a single dominant taxon is highly abundant (e.g., blackflies abundant in large numbers), other taxa may not be detected because they account for a very low fraction of organisms.

A fixed count of 100 organisms from each composite sample is expected to detect taxa that constitute 2% or more of the organisms in the composite sample with over 87% probability (Table 3-5). With increasing relative abundance of a taxon, the probability of detection rapidly approaches 100%. For rare taxa (< 1% of the organisms in the composite) the probability of detection is 63% or less for a subsample of 100 organisms, and 87% for a subsample of 200 organisms (Table 3-5). A subsample of 230 organisms or more is required to detect rare taxa with 90% probability (Table 3-6).

Table 3-5. Probability (%) of detecting at least one organism of a taxon X with relative abundance P for varying subsample sizes n.

	Density P as fraction (%) of organisms in the entire composite sample				
n	1	2	3	4	5
100	63	87	95	98	≈ 100
200	87	98	≈ 100	≈ 100	≈ 100
300	95	≈ 100	≈ 100	≈ 100	≈ 100
400	98	≈ 100	≈ 100	≈ 100	≈ 100
500	99	≈ 100	≈ 100	≈ 100	≈ 100

Table 3-6. Subsample sizes n required to achieve at least 90% probability of detecting a taxon *X* that constitutes a proportion P of the composite sample.

	Density P as fraction (%) of organisms in the entire composite sample				
	1	2	3	4	5
n	230	114	76	57	45

The number of plots sampled in each stream segment, and the number of organisms subsampled for identification in the lab, affects the precision in estimates of species richness and in particular the likelihood of detecting rare species. Underestimation of rare species can reduce the sensitivity of community-based assessment methods to detect ecological changes, and thus reduce the effectiveness of bioassessment (Cao et al. 1998). Cao et al. (1998) advocate a larger sample size than the standard of 100 to 300 individuals used in EPA rapid bioasssessment protocols (RBPs) to reliably differentiate between reference and impacted sites. However, the actual count of organisms required to achieve adequate power for distinguishing between reference and impaired sites depends on characteristics of the biota and thus may differ among studies. Somers et al. (1998), for example, compared biological indices for assessing health of lakes based on counts of 100, 200, and 300 organisms, and found that doubled or tripled effort resulted in little improvement in the ability to distinguish between lakes.

3.7.1.5 Recommendation for Future Benthic Field Study

We recommend that a field experiment be conducted to address the multiple factors that may affect BIBI comparability. This experiment should employ a study design and sample size that facilitates analysis of the issues discussed in this chapter. Ideally, sampling would take place at randomly selected sites that represent the full range of conditions. At present, plans are for such an experiment to be conducted by MBSS and Montgomery County DEP. Subsequent data analysis would provide many answers to IBI calibration issues; this is necessary before full program integration can be implemented. If funding can be obtained and appropriate site locations identified, the study would be conducted in conjunction with planned sampling by both MBSS and Montgomery County in Seneca Creek watershed in Spring 2001. The fact that both programs are scheduled to sample in Seneca Creek watershed at the same time is fortuitous and provides the ideal opportunity for conducting this field experiment. In Section 3.8, we describe details of a proposed pilot study design.

3.7.2 FIBI Comparability

3.7.2.1 Analysis of MBSS data

To isolate differences resulting from data analysis and IBI conventions (while controlling for field method), MBSS fish data were used to compare the MBSS FIBI and Montgomery County FIBI. FIBIs were calculated using raw data from 54 sites sampled by MBSS in Montgomery County during the 1995-1997 Survey. Both the numerical IBI scores and narrative IBI ratings were compared.

The Montgomery County FIBI yielded results similar to the MBSS FIBI for the MBSS sites analyzed (Figure 3-16). Categorical ratings were quite similar, with discrepancies most common at higher quality sites (MBSS good-to-fair or Montgomery County excellent-to-good sites). Differences appear to be attributable to the use of different metrics, because protocols for counting and identifying fish are the same across programs. Comparability between the two FIBIs was stronger than between BIBIs in the parallel example presented above, perhaps because FIBI metrics used by the two programs are more similar than the BIBI metrics and because the field methods that these metrics are based on differ only slightly. Specifically, Montgomery County has in the past used three electrofishing passes, while MBSS uses two passes.

3.7.2.2 Analysis of Montgomery County data

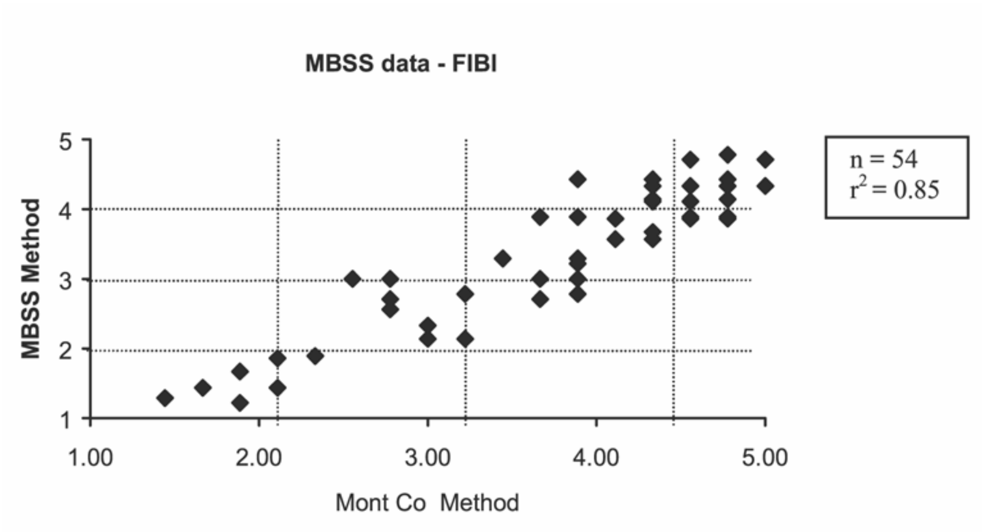
As an additional test to isolate differences resulting from data analysis and IBI conventions while controlling for field methods, Montgomery County fish data were used to compare the MBSS FIBI and Montgomery County FIBI. FIBIs were calculated using raw data from 237 sites sampled by Montgomery County during 1995-1999. As above, IBI scores were compared via scatter plots and linear regression; IBI narrative ratings were compared using categorical analysis.

While some sites tended to receive similar ratings, some inconsistencies were apparent (Figure 3-17). More differences were observed between MBSS and Montgomery County IBI scores in this analysis than were found with the MBSS data set (Figure 3-16); reasons for this difference were not clear. Montgomery County data from the third pass was used in this analysis; we recognize that the use of three- vs. two-pass data may have had a slight effect on the result (although this effect should be only minimal, as shown in 3.7.2.4 below). Remaining differences are more likely the result of the different metrics and thresholds employed in the two IBI formulations.

3.7.2.3 Analysis of 1997 MBSS-Montgomery County Joint Field Study Data

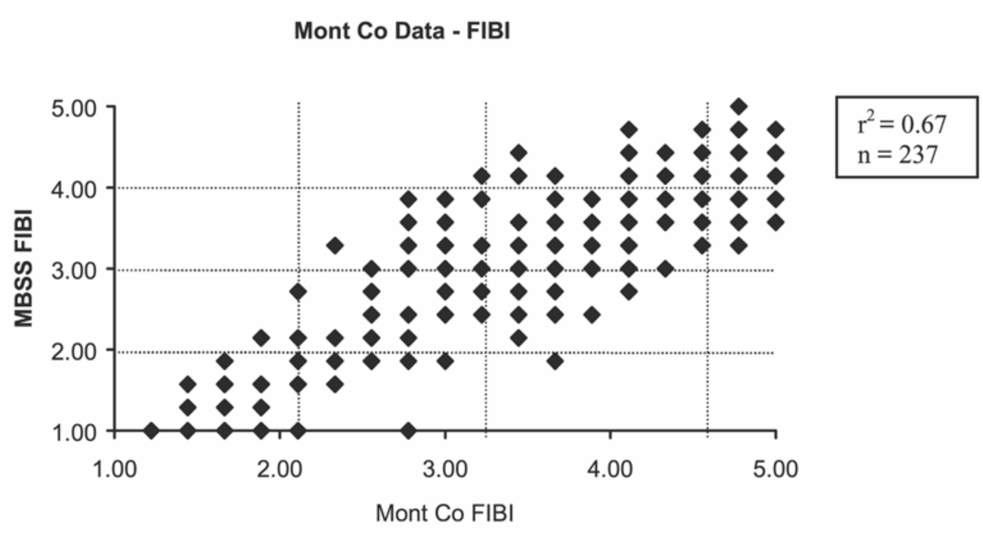
During the 1997 joint sampling study described in Section 3.7.1.3, fish data were collected by joint MBSS-Montgomery County field crews at nine sites; two other reaches contained sites that by chance were sampled for fish by both programs. Data from 10 of these sites were available for analysis.

Data were scored using the MBSS and Montgomery County FIBIs. A comparison shows some promising results, but data were limited to very few sites, all with little or no degradation (Figure 3-18). Also, narrative ratings varied. MBSS rated five sites in its top narrative category, while Montgomery IBI rated no sites as excellent; note that the threshold for a County rating of "excellent" is 4.5 while the MBSS threshold for its top rating of "good" is 4.0. Consistent scoring thresholds for narrative categories would be desirable in program integration, particularly if ratings lead to specific management strategies (e.g., identification of high quality areas for conservation). A more complete evaluation of FIBI comparability would include data from a broad range of conditions. In this joint study, field crews from both programs worked together, producing a single set of data for each shared site. If crews worked separately, taxonomic identification accuracy could potentially differ between field crews, resulting in differences in IBI scores when sampling the same fish assemblage.



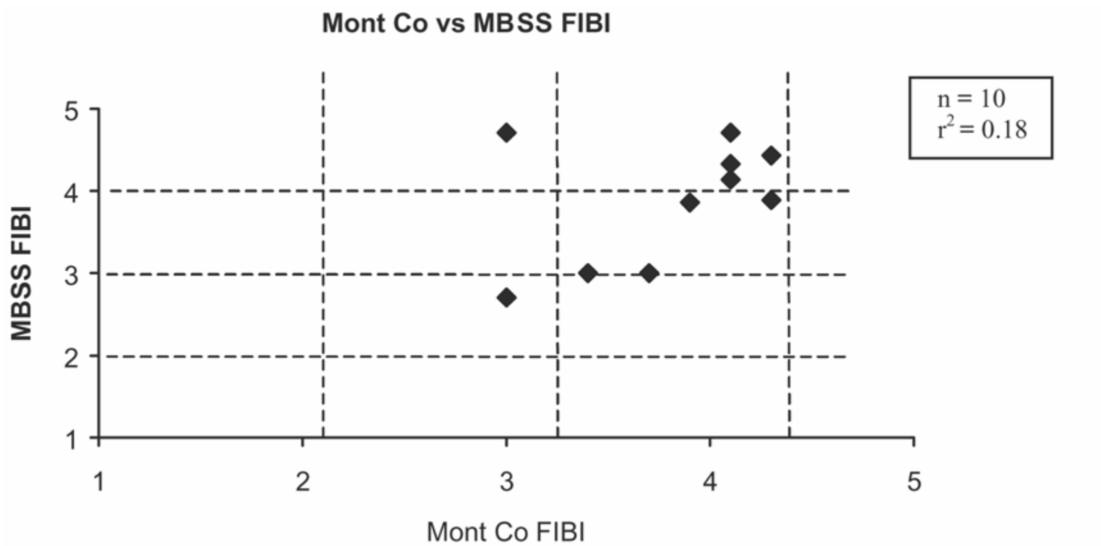
FIBI – MBSS Method	Good	0	0	**8**	**9**
	Fair	0	2	**12**	**4**
	Poor	0	**6**	4	0
	Very Poor	**8**	1	0	0
		Poor	Fair	Good	Excellent
	FIBI - Montgomery County Methods				

Figure 3-16. Comparisons of MBSS and Montgomery County FIBI numeric scores and narrative ratings (each program uses different naming conventions for the ratings but they are comparable as aligned in the table), using MBSS data.



FIBI – MBSS Method					
	Good	0	14	**1**	**24**
	Fair	0	27	**56**	**47**
	Poor	4	**16**	15	0
	Very Poor	**23**	9	1	0
		Poor	Fair	Good	Excellent
	FIBI - Montgomery County Methods				

Figure 3-17. Comparisons of MBSS and Montgomery County FIBI numeric scores and narrative ratings (each program uses different naming conventions for the ratings but they are comparable as aligned in the table), using Montgomery County data.



FIBI – MBSS Method					
	Good	0	0	**5**	**0**
	Fair	0	0	**4**	**0**
	Poor	0	**1**	0	0
	Very Poor	**0**	0	0	0
		Poor	Fair	Good	Excellent
	FIBI - Montgomery County Methods				

Figure 3-18. Comparison of Montgomery County and MBSS FIBI numeric scores, 1997 joint sampling study (each program uses different naming conventions for the ratings but they are comparable as aligned in the table).

3.7.2.4 Fish Abundance and IBI Scores Based On Two vs. Three Electrofishing Passes

Because Montgomery County uses three-pass electrofishing, the County's data allow examination of the effect of two vs. three electrofishing passes on fish metrics and IBI scores. Electrofishing data from 322 countywide sites sampled by Montgomery County were analyzed. Within this data set, the total number of species was equal using either two or three passes (i.e., no new species were collected on the third pass at any site). Fish abundance (total number of fish captured) did increase slightly with three passes. However, the total number of fish captured with two passes was highly correlated with the total from three passes (Figure 3-19). Observed effects on FIBI results were minor. FIBI scores calculated from two passes were highly correlated with FIBIs based on three passes and ratings by category were nearly unchanged (Figure 3-20).

3.7.2.5 Recommendation for Integration of FIBI

Because of the data already available in this and other studies to compare the two-pass vs. three-pass electrofishing methods, and because field protocols are otherwise identical (i.e., both identify all fish captured in multiple passes of a 75-m reach), we do not find that additional field studies are needed to compare fish field protocols used by MBSS and Montgomery Counties. In general, if two programs had concerns about whether taxonomic identifications were accurate and consistent across programs, further study would be recommended.

To integrate fish bioassessment results, fish data can be shared among the MBSS and Montgomery County programs. Further coordination between programs on future IBI refinements could make IBI results from the two programs more consistent than at present, (i.e., for joint assessments and reporting), although current information on FIBI consistency is encouraging. If each program wishes to retain its own FIBI for its own use, the following recommendations would apply:

> MBSS - use two-pass Montgomery County data (i.e., dropping data from third pass) and calculate MBSS FIBI.
>
> Montgomery County - if two-pass MBSS data are used, fish abundance measures could be calculated by extrapolation from two to three passes. The observed effect of third pass appeared to have a minimal effect on IBI scores.

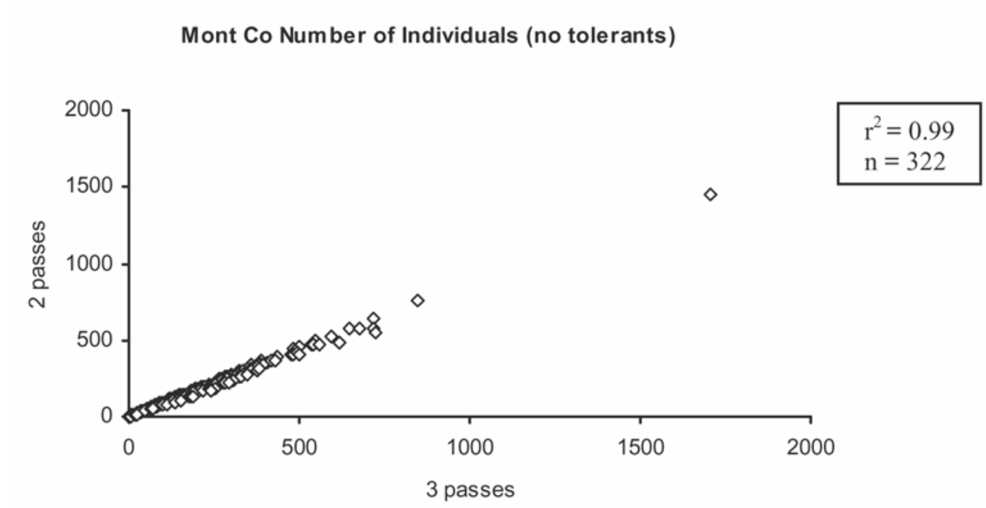


Figure 3-19. Montgomery County metric for total number of fish (excluding tolerants), comparing values from two vs. three electrofishing passes, using Montgomery County data.

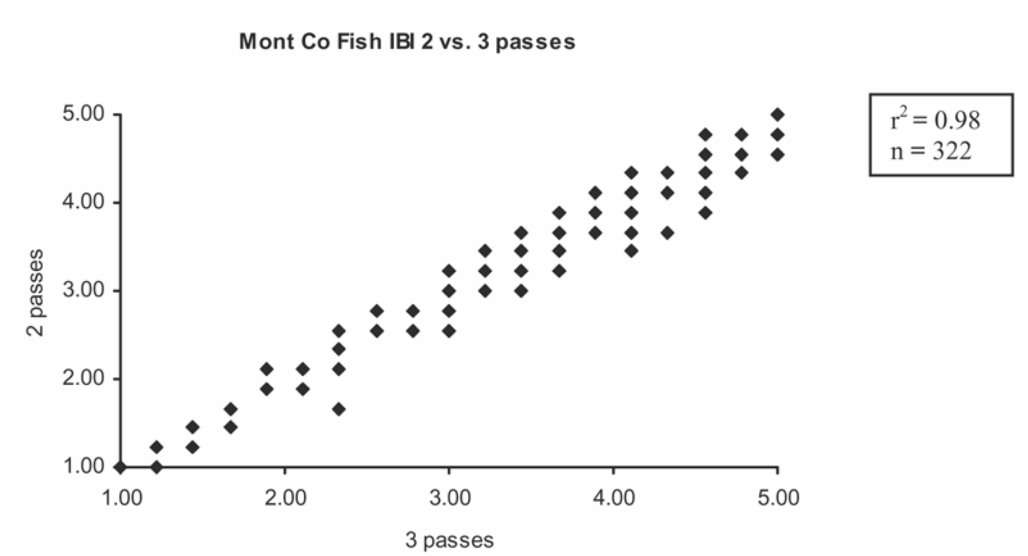


Figure 3-20. Comparison of FIBI scores from two vs. three electrofishing passes, using Montgomery County data.

3.8 OPTIONS FOR COMBINING PROGRAM RESULTS

3.8.1 Developing An Integrated Approach To Estimating Stream Condition

If areawide estimation of stream conditions using integrated data is a goal (as it is for MBSS), integration requires development of an analytical approach that allows estimation of parameters of interest (means, percentage of stream miles) with quantifiable errors (known precision). Several options exist for integrating data from multiple programs:

(1) The simplest and most effective option would be for both programs to use a consistent, unified study design that would lead to straightforward and cost-effective integration, providing a greater number of samples and increased precision in estimates for a local area. However, this approach is not feasible where programs need to maintain sites or site selection procedures already in use (such as integration of MBSS and Montgomery County).

(2) Maintaining separate study designs and using a joint estimation of stream condition is possible, but requires more complex analysis. This analysis requires variance estimates for both studies (note that variance estimates are also essential for the State's biocriteria listing framework). Estimates may be combined into a single estimate via weighting procedures. Requirements for joint estimation by weighting include:

Analyzing data for each program, consistent with individual survey designs;

Estimating means and variances for each program;

If variance estimates not already available, gathering ancillary data on the number of reaches and/or stream miles in each level/subdivision of the sample frame (by stream order) and calculating properly weighted estimates and variances (minimum two PSUs per stratum); and

Assuming that all stream miles are able to be sampled.

In our case study, this integrated analysis requires detailed data on the number of stream reaches and stream miles in each subwatershed unit. Data need to reflect the unique, overlapping, and total stream miles in both sample frames, broken down by stream order. An analytical approach for the MBSS-Montgomery County integration is presented in Section 3.8.2 below. Note that this analysis must be tailored to each sampling program, a process that could be resource intensive (e.g., other Maryland counties that want to integrate with MBSS would require new analysis).

(3) A third option would be to maintain current county and state designs, but coordinate the site selection (e.g., replacing MBSS site with Montgomery County site when nearby). This would provide some savings by avoiding duplication of field effort, but would require complex methods to create a joint estimate using all the sites from both programs. Field protocol and indicator consistency would still need to be resolved before data from shared sites could be used by both programs. If field methods are consistent, both programs could use the shared sites in their standard estimation procedures to produce separate county and state estimates, if desired.

If procedures and indicators differ, special field studies and analysis are needed to demonstrate field protocol and indicator equivalency. Only after demonstrating equivalency would it be appropriate to propose site replacement. For example, MBSS and Montgomery County will collect full data for both programs during the 2001 sampling season, and at the same time conduct a pilot study on data comparability. Results may support future site replacements. For example, a joint site selection approach for MBSS and Montgomery County might include the following provisions, after both programs have made preliminary site selections in the same watershed:

If an MBSS site falls on a reach with no Montgomery County site, add the MBSS site.

If an MBSS site falls on the same reach and is close to a Montgomery County site, use either Montgomery County site or MBSS site in both programs.

If an MBSS site falls on the same reach but is far from Montgomery County site, add the MBSS site and use it to evaluate the length of the reach with same condition.

A review of existing data (Figure 3-21) suggests that a distance cutoff of 500 meters would be an appropriate distinction between “close” and “far” for this procedure. Note that site replacement could only practically occur at sites that are targeted to be sampled in the same year.

Finally, if derivation of joint, areawide estimates is not a goal in program integration, or is not deemed possible, other options for joint assessments include the use of county/local data to improve the understanding of the causes of impairments within small watersheds (e.g., MD 8-digit or 12-digit). State programs such as MBSS typically cannot conduct sampling at a high density within all local areas. County and local data (from both targeted and random sites) could be incorporated into the targeted component of MBSS sampling, when information is needed about particular problem streams or high-quality systems.

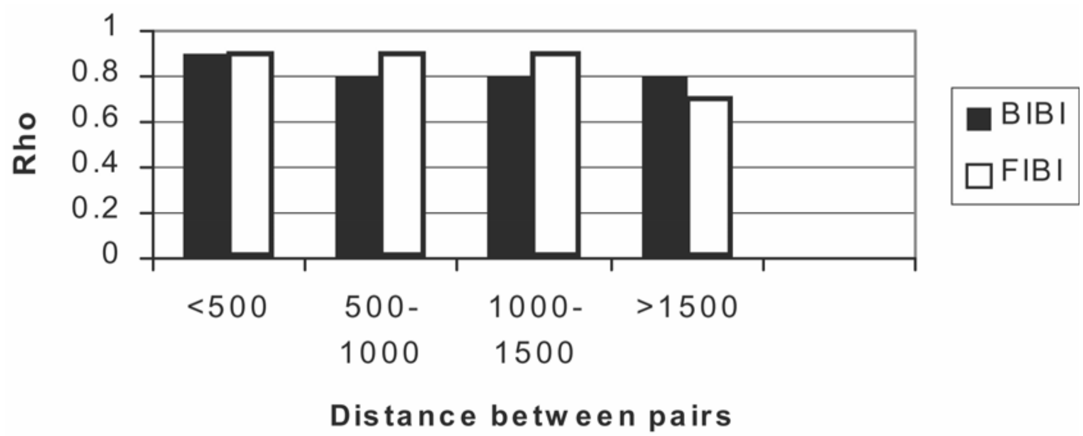


Figure 3-21. Variability in IBI scores at sites within the same reach, by distance between sites.

3.8.2 Integration Example: Analytical Approach for MBSS and Montgomery County Surveys

The primary objective of combining Montgomery County and MBSS surveys is to obtain a unified estimate of stream condition with less variance than the individual estimates. Several differences between MBSS and Montgomery County surveys must be accounted for in the analysis. The two programs use different sampling frames, with different spatial coverage. Map studies show that the streams on the 1:100,000 map used in MBSS can be treated effectively as a subset of the streams on the 1:24,000 map used by Montgomery County. The coarser scale map used in MBSS primarily misses some of the smaller headwater streams. In deriving a joint estimate, the streams that are exclusive to the Montgomery County sampling frame can be considered a separate stratum, with estimates of stream condition for this stratum based on Montgomery County data. For the streams appearing on both the 1:24,000 and 1:100,000 maps (overlapping area of the sample frames), we estimated the watershed mean for the parameter of interest (e.g., mean BIBI) using a composite estimation technique (Korn and Graubard 1999, p. 282) that combines the individual survey weighted means,

$$\bar{y} = \frac{\left(k_1 \sum_{i=1}^{n_1} w_{i1}\right)\bar{y}_1 + \left(k_2 \sum_{i=1}^{n_2} w_{i2}\right)\bar{y}_2}{\left(k_1 \sum_{i=1}^{n_1} w_{i1}\right) + \left(k_2 \sum_{i=1}^{n_2} w_{i2}\right)}$$

where

$\bar{y}_1$ and $\bar{y}_2$ are the estimated means for Montgomery County and MBSS, respectively;

the values of k_1 and k_2 determine how much each survey mean contributes to the combined mean;

w_{ij} is the weight for sample *i* in survey *j* (*j*= 1,2); and

n_1 and n_2 are the sample sizes for Montgomery County and MBSS.

The sum of the sample weights for each survey sums up to the total population size covered in the survey area (restricted here to streams within the map overlay area), as

$N_1 = \sum_{i=1}^{n_1} w_{i1}$ = number of stream segments in the overlap area for Montgomery County survey

$N_2 = \sum_{i=1}^{n_2} w_{i2}$ = number of stream segments in the overlap area for MBSS.

The total number of stream segments in the survey area is estimated by dividing the total number of stream miles in the overlap area by the stream segment length. Accurate estimates of the stream length are available from GIS analysis of the maps. For simplicity, we assume that the number of stream miles and, hence, segments is equal for the two surveys, or $N_1 = N_2 = N$. The combined mean thus refers to one map scale and can be estimated by the simplified formula

$$\bar{y} = \frac{k_1\bar{y}_1 + k_2\bar{y}_2}{k_1 + k_2}$$

It is desirable to assign the highest influence to the most precise survey estimate. To minimize the variance of the combined estimate $\bar{y}$, set

$$k_j = \left(1 - \frac{Var(\bar{y}_j)}{Var(\bar{y}_1) + Var(\bar{y}_2)}\right)$$

Only probability-based samples can be used in this combined estimate, because estimates of the variances are required to determine the weights. For Montgomery County, sites that are targeted will not be included in this estimation because their inclusion probability is unknown.

Variance estimation

The selection of primary sampling units (PSUs) is performed independently in the MBSS and Montgomery County surveys. The PSUs are defined differently in the two surveys and the stratification is different. Assume that Montgomery County employed L_1 strata and MBSS L_2 strata. For variance estimation, we treat the population of stream segments as if it was composed of $L = L_1 + L_2$ strata. This stratification controls for survey differences (Korn and Graubard 1999). Because the two surveys are independent,

$$Var(\bar{y}) = \sum_{i=1}^{L} w_i^2 Var(\bar{y}_i)$$

where the strata weights are

$$w_i = \frac{N_i}{\sum_{i=1}^{L} N_i}.$$

This estimator can be expanded to include the Montgomery County estimate for streams that are not covered by the 1:100,000 map. An example of applying this methodology for Seneca Creek is presented in a separate manuscript (Vølstad et al. in prep.).

3.8.3 Seneca Creek Pilot Study

MBSS and Montgomery County are hoping to conduct a pilot study in Seneca Creek watershed during 2001. One primary purpose of the pilot study is to calibrate IBIs, so that data from the two programs may be combined to derive a single unified statement about stream conditions. Calibration would also support future reductions in field effort by allowing site replacement (i.e., one program using data collected by the other program, rather than both sampling the same or a nearby site).

A preliminary list of specific questions to be answered by this study was prepared and will be refined through further discussions with EPA, MBSS, and Montgomery County program managers. Proposed questions include:

> How comparable are MBSS and Montgomery County IBIs? Do they give similar ratings of stream condition?
>
> What is the variability in Montgomery County IBI scores? MBSS IBI scores? Variability between programs?

> What is the effect of the number of plots (20 D-net vs. 2 kick net samples) on benthic assessments?
>
> What is the effect of taking 100 vs. 200 organism subsamples on benthic bioassessment results?
>
> What effect does identifying oligochaetes and chironomids to genus level, compared with higher-level taxonomy, have on benthic bioassessment results?
>
> Can estimates of stream condition in Seneca Creek Watershed be improved (i.e., made more precise) by integrating Montgomery County data with MBSS?

In addition, laboratory differences in subsampling and taxonomic identification accuracy may be investigated.

3.8.4 Proposed Pilot Study Design

The primary objective of this pilot study is to evaluate whether Montgomery County and MBSS field sampling protocols for benthos result in the same classification of stream condition. Replicate sampling was conducted by Montgomery County and MBSS in 1997 at 12 sites for benthos. Although scores from the two programs generally were in the same (or neighboring) categories, the results were inconclusive because the sites had little variation in IBI scores.

We propose to conduct an experiment that compares benthic sampling protocols under a variety of stream conditions. The experimental design will test whether the difference between BIBI scores from replicate samples (within stream segments) collected by the MBSS and Montgomery County is significantly larger than the expected differences for the same field protocol. We propose to conduct replicate sampling between programs, and within programs, using an experimental design that is effective for detecting the effects of sampling protocol on IBI scores.

Using the statistical terminology of experimental designs, the field sampling method can be seen as a "treatment," the results of the treatments being the differences in IBI scores between replicates.

The proposed experiment involves three treatments, described in Table 3-7. We propose a two-way replicated factorial design (Box et al. 1978). The three treatments will be tested in four groups of streams, with 4 replications (Table 3-8).

We propose this design to ensure that sampling protocol comparisons are conducted under varying stream conditions. The "replications" involve random selection of 4 stream segments in each combination. This design allows us to test effects of sampling protocol and stream type on differences in IBI scores for replicate sampling. It also allows us to test if stream type influences the differences in IBI scores between replicate samples (e.g., are Montgomery County and MBSS scores more similar in small streams than in large streams).

Table 3-7. Proposed treatments for pilot study.

Treatment	Explanation
A	MBSS – Montgomery Co. replicate sampling within stream segment
B	Montgomery Co. – Montgomery Co. replicate sampling within stream segment
C	MBSS - MBSS replicate sampling within stream segment

Table 3-8. Proposed design for pilot study.

Stream Type		**Treatment**			**Number of Sites per Block**
% Urban	Stream order (based on 1:100K map)	A	B	C	
High	1,2	4	4	4	12
	3+	4	4	4	12
Low	1,2	4	4	4	12
	3+	4	4	4	12
# sites per treatment		16	16	16	48 (total)

As part of this comparison study, we also propose to evaluate the effects of benthic subsample size in the laboratory and taxonomic classification level (of oligochaetes and chironomids) on IBI scores. To evaluate the effects of subsample size, two subsamples of 100 organisms each would be identified separately in the laboratory; results would then be analyzed separately or grouped, as needed for comparisons. To evaluate taxonomic classification, oligochaetes and chironomids would be mounted and identified to genus; results could then be analyzed at genus level or grouped, as needed. Options for making these comparisons are described below:

Evaluate effects of subsampling 100 vs. 200 organisms

Option I: 100+100 organisms at all (48) stations; keeping results separate for the two groups of 100;

Option II: 100+100 organisms for treatment A (16 stations).

Evaluate effect of taxonomic classification

Option I: Montgomery County classifies oligochaetes and chironomids to genus level at all 32 station in treatments A+B;

Option II: Montgomery County classifies oligochaetes and chironomids to genus level at 16 stations for treatment A.

The first of these evaluations primarily involves extra effort by the MBSS program; while the second part involves more effort for Montgomery County.
The data will be analyzed using ANOVA to determine the effects of sampling protocol and stream type on the difference in IBI scores for replicate sampling. The mathematical model for this experiment can be written as (Hicks 1993, p. 129):

$$Y_{ijk} = \mu + T_i + S_j + TS_{ij} + \varepsilon_{k(ij)}$$

where T_i represents the sampling method effect, S_j the stream condition effect (% urban and stream order), and TS_{ij} the interaction between stream condition and sampling method; i = 1,2,3 for the three field comparison types; j = 1,2,3,4 for the four classes of streams; and k is the number of observations for each i, j combination.

Based on this ANOVA model, we can test the following hypotheses:

$$H_{01} = T_i = 0$$
$$H_{02} = S_j = 0$$
$$H_{03} = TS_{ij} = 0$$

for all *i* and *j*. This experimental design is effective for detecting differences between sampling protocols because the comparisons are done under different stream conditions. The testing of interaction is also important, in case differences in methods performance relates to stream condition (e.g., resulting in more similar scores for streams with poor IBI scores in urban areas).

The IBI scores based on 100 or 200 organisms, and for the two taxonomic classification levels, will be analyzed to test if increased sub-sample size and classification to genus (for oligochaetes and chironomids) significantly reduces the difference between Montgomery County and MBSS IBI scores.

We will use the same analytical techniques employed in the variability study (Section 2) to compare the programs.

An added benefit of the proposed design is that we can test whether variability in IBI scores at a site tends to increase with stream order. Results of the variability analysis based on MBSS data indicated that replicate IBI scores are more variable for third-order streams than for lower stream orders. Also, replicates at impaired sites tended to be more similar than replicates at sites with good scores. The proposed experiment would provide further information to determine how uncertainty in IBI scores relates to stream condition. High and low % urban land use would serve as a proxy for impaired and non-impaired streams (Roth et al. 1998).

4. REFERENCES

Agresti, A. (1990). *Categorical Data Analysis*. John Wiley & Sons. New York.

Barbour, M.T. and J.Gerritsen. (1996). Subsampling of benthic samples: A defense of the fixed-count method. *J.N. Am. Benthol. Soc.* **15**: 386-391.

Boswell, M.T., Burnham, K.P. and Patil, G.P. (1988). Role and use of composite sampling and capture-recapture sampling in ecological studies. In: P.R. Krishnaiah and C.R. Rao, eds., *Handbook of Statistics, Vol. 6*. Elsevier Science Publishers.

Box, G.E.P., Hunter, W. and Hunter, J.S. (1978). *Statistics for Experimenters*. John Wiley & Sons. New York.

Cao, Y., Williams, D.D., and Williams, N.E. (1998). How important are rare species in aquatic community ecology and bioassessment? *Limnol. Oceanogr.* 43(7): 1403-1409.

Cochran, W.G. (1977). *Sampling Techniques.* 3rd Edition. John Wiley and Sons, New York.

Courtemanch, D.L. (1996). Commentary on the subsampling procedures used for rapid bioassessments. *J. N. Am. Benthol. Soc.* 15: 381-385.

Cronbach, L.J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrica* 16: 297-334.

Ebel, R.L. (1951). Estimation of the reliability of ratings. *Psychometrika* 16: 407-424.

Edland, S.D. and van Belle, G. (1994). Decreased sampling costs and improved accuracy with composite sampling. In: Cothern, C.R. and N.P. Ross, eds., *Environmental Statistics, Assessment, and Forecasting.* Lewis Publishers.

Elliott, J.M. (1971). *Some Methods for the Statistical Analysis of Samples of Benthic Invertebrates.* Freshwater Biological Association. Scientific Publication No. 25.

Gilbert, R.O. (1987). *Statistical Methods for Environmental Pollution Monitoring.* Van Nostrand Reinhold. New York.

Haggard, E.A. (1958). *Intraclass correlation and the analysis of variance*. NY: Dryden.

Heimbuch, D.G., Seibel, J.C., Wilson, H.T. and Kazyak, P.F. (1999). A multiyear lattice sampling design for Maryland-wide fish abundance estimation. *J. Agr., Biol. and Env. Stat.* **4**: 443-455.

Hicks, C.R. (1994). *Fundamental Concepts in the Design of Experiments*. 4th Edition. Saunders College Publishing. New York.

Hughes, R.M., Kaufmann, P.R., Herlihy, A.T., Kincaid, T.M., Reynolds, L. and Larsen, D.P. (1998). A process for developing and evaluating indices of fish assemblage integrity. *Can. J. Fish. Aquat. Sci.* 55: 1618-1631.

Jessen, R.F. (1978). *Statistical Survey Techniques.* John Wiley & Sons. New York.

Kazyak, P.F. (2000). Maryland Biological Stream Survey Sampling Manual. Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division, Annapolis.

Klauda, R., Kazyak, P., Stranko, S., Southerland, M., Roth, N., and Chaillou, J. (1998). Maryland Biological Stream Survey: A state agency program to assess the impact of anthropogenic stresses on stream habitat quality and biota. *Environmental Monitoring and Assessment* 51: 299-316.

Korn, E.L. and Graubard, B.I. (1999). *Analysis of Health Surveys*. John Wiley & Sons. New York. 382.

Litwin, M.S.(1995). *How to Measure Survey Reliability and Validity*. Survey Kit series, Vol. 7. Thousand Oaks, CA: Sage Publications. ISBN: 0803957041.

Maryland Department of Environment (MDE). 2000. Report of the Biological Criteria Advisory Committee to the Maryland Department of the Environment on the Interim Framework for the Regulatory Application of Biological Assessments. R. Eskin, J.E. Lathrop-Davis, and T.C. Rule, eds.

May, R.M. 1975. Patterns of species abundance and diversity. In: M.L. Cody and J.M. Diamond (editors). *Ecology and evolution of communities.* Harvard University Press, Cambridge, Massachusetts. pp. 81-120.

McCammon, B.P. (1994). Recommend watershed terminology. *In:* Watershed Management Council newsletter, Vol. 6 No. 2, Fall 1994. **www.watershed.org.**

Mercurio, G., Chaillou, J.C., and Roth, N.E. (1999). Guide to Using 1995-1997 Maryland Biological Stream Survey Data. Prepared by Versar, Inc., Columbia, MD, for Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division. CWBP-MANTA-EA-99-5.

Patil, G.P., Gore, S.D. and Sinha, A.K. Environmental chemistry, statistical modeling, and observational economy. In: Cothern, C. R. and N. P. Ross, eds., *Environmental Statistics, Assessment, and Forecasting.* Lewis Publishers.

Roth, N.E., Southerland, M.T., Mercurio, G., Chaillou, J.C., Heimbuch, D.G. and Seibel, J.C. (1999). State of the Streams: 1995-1997 Maryland Biological Stream Survey Results. Prepared by Versar, Inc., Columbia, MD, and Post, Buckley, Schuh, and Jernigan, Bowie, MD, for Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division. CBWP-MANTA-EA-99-6.

Roth, N.E., Southerland, M.T., Chaillou, J.C., Kazyak, P.F. and Stranko, S.A. (2000). Refinement and Validation of a Fish Index of Biotic Integrity for Maryland Streams. Prepared by Versar, Inc., Columbia, MD, with Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division.

Roth, N., Southerland, M., Chaillou, J., Klauda, R., Kazyak, P., Stranko, S., Weisberg, S., Hall, Jr., L. and Morgan II, R. (1998a). Maryland Biological Stream Survey: Development of a fish index of biological integrity. *Environmental Monitoring and Assessment* 51: 89-106.

Roth, N.E., Southerland, M.T., Strebel, D.E. and Brindley, A. (1998b). Landscape Model of Cumulative Impacts: Phase I Report. Prepared by Versar, Inc., Columbia, MD, for Maryland Department of Natural Resources.

Seber, G.A.F. (1973). *The Estimation of Animal Abundance.* Griffin. London. 506.

Snedecor, G.W. and Cochran, W.G. (1980). *Statistical Methods.* The Iowa State University Press.

Somers, K.M., Reid, R.A. and David, S.M. (1998). Rapid biological assessments: how many animals is enough? *J. N. Am. Benthol. Soc.* **17**: 348-358.

Southerland, M., Roth, N., Mercurio, G. and Vølstad, J. (2000). Final Design and Procedures for MBSS 2000-2004 (Round Two). Memorandum to R. Klauda and P. Kazyak, Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division. February 10, 2000.

Strahler, A.N. (1957). Quantitative analysis of watershed geomorphology. *Transactions American Geophysical Union*, vol. 38, pp. 913-920.

Stribling, J.B., Jessup, B.K., White, J.S., Boward, D. and Hurd, M. (1998). Development of a Benthic Index of Biotic Integrity for Maryland Streams. Prepared by Tetra Tech, Inc., Owings Mills, MD and Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division. CBWP-MANTA-EA-98-3.

Van Ness, K., Brown, K., Haddaway, M.S., Marshall, D. and Jordahl, D. (1997). Montgomery County Water Quality Monitoring Program Stream Monitoring Protocols. Montgomery County Department of Environmental Protection, Watershed Management Division, Rockville, MD. February 20, 1997.

Vølstad, J.H., Southerland, M.T., Weisberg, S.B., Wilson, H.T., Heimbuch, D.G. and Seibel, J.C. (1996). *Maryland Biological Stream Survey: The 1994 Demonstration Project*. Report prepared by Versar, Inc., Columbia, MD, for Maryland Department of Natural Resources, Monitoring and Non-Tidal Assessment Division, MD. CBWP-MANTA-EA-95-9.

Vølstad, J.H., Southerland, M., Chaillou, J., Wilson, H.T., Heimbuch, D.G., Jacobson, P.T. and Weisberg, S.B. (1995). *Maryland Biological Stream Survey: The 1993 Pilot Study*. Report prepared by Versar, Inc., Columbia, MD, for Maryland Department of Natural Resources, Chesapeake Bay Research and Monitoring Division, Annapolis, MD. CBRM-AD-95-3.

Vølstad, J.H., Neerchal, N.K., Roth, N.E. and Southerland, M.T. *In preparation*. Combining Biotic Indices of Stream Condition from Multiple Surveys in a Maryland Watershed.